

NASA/TM-2002-211444



# Research on Hazardous States of Awareness and Physiological Factors in Aerospace Operations

*Lawrence J. Prinzel III*  
*Langley Research Center, Hampton, Virginia*

---

March 2002

## The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at <http://www.sti.nasa.gov>
- Email your question via the Internet to [help@sti.nasa.gov](mailto:help@sti.nasa.gov)
- Fax your question to the NASA STI Help Desk at (301) 621-0134
- Telephone the NASA STI Help Desk at (301) 621-0390
- Write to:  
NASA STI Help Desk  
NASA Center for AeroSpace Information  
7121 Standard Drive  
Hanover, MD 21076-1320

NASA/TM-2002-211444



# Research on Hazardous States of Awareness and Physiological Factors in Aerospace Operations

*Lawrence J. Prinzel III*  
*Langley Research Center, Hampton, Virginia*

National Aeronautics and  
Space Administration

Langley Research Center  
Hampton, Virginia 23681-2199

---

March 2002

---

Available from:

NASA Center for AeroSpace Information (CASI)  
7121 Standard Drive  
Hanover, MD 21076-1320  
(301) 621-0390

National Technical Information Service (NTIS)  
5285 Port Royal Road  
Springfield, VA 22161-2171  
(703) 605-6000

## ABSTRACT

The technical memorandum describes research conducted to examine the etiologies and nature of hazardous states of awareness and the psychophysiological factors involved in their onset in aerospace operations. A considerable amount of research has been conducted at NASA that examines psychological and human factors issues that may play a role in aviation safety. The technical memorandum describes some of the research that was conducted between 1998 and 2001, both in-house and as cooperative agreements, which addressed some of these issues. The research was sponsored as part of the physiological factors subelement of the Aviation Operation Systems (AOS) program and Physiological / Psychological Stressors and Factors project. Dr. Lance Prinzel is the Level III subelement lead and can be contacted at [l.j.prinzel@larc.nasa.gov](mailto:l.j.prinzel@larc.nasa.gov)



## TABLE OF CONTENTS

Abstract . . . . .	1
Introduction . . . . .	4
Recent Physiological Factors Research. . . . .	8
Hazardous States of Awareness . . . . .	8
Adaptive Automation . . . . .	22
Individual Difference Variate Research . . . . .	46
Training for Hazardous States of Awareness Research . . . . .	51
Spin-Off Research . . . . .	63
Current Research Focus . . . . .	69
References . . . . .	77

## INTRODUCTION

The purpose of the NASA Technical Memorandum is to describe some of the issues surrounding the use of automation in complex systems, its effect on the human operator, and recent research that has been conducted at NASA that has addressed some of these issues. For additional details about any of the studies described in the document, please contact Dr. Lance Prinzl at NASA Langley Research Center ([l.j.prinzel@larc.nasa.gov](mailto:l.j.prinzel@larc.nasa.gov)).

### Automation

Automation refers to "...systems or methods in which many of the processes of production are automatically performed or controlled by autonomous machines or electronic devices" (p.7). Automation is a tool, or resource, that the human operator can use to perform some task that would be difficult or impossible without the help of machines (Billings, 1997). Therefore, automation can be thought of as a process of substituting some device or machine for some human activity; or it can be thought of as a state of technological development (Parsons, 1985). However, some people (e.g., Woods, 1996) have questioned whether automation should be viewed as a substitution of one agent for another. Nevertheless, the presence of automation has pervaded every aspect of modern life. We have built machines and systems that not only make work easier, more efficient and safer, but also have given us more leisure time. The advent of automation has further enabled us to achieve these ends. With automation, machines can now perform many of the activities that we once had to do. Now, automatic doors open for us. Thermostats regulate the temperature in our homes for us. Automobile transmissions shift gears for us. We just have to turn the automation on and off. One day, however, there may not be a need for us to do even that.

The evolution of technology has generated modern machines and systems that expand the range of human capabilities enormously. The benefits afforded by such technological progress necessitate machines and systems of greater sophistication and complexity. Consequently, the demands placed upon the operators of these systems increase along with the growth in complexity. The introduction of automation into systems has helped operators manage the complexity, but has not necessarily relieved the burden of interacting with such systems. Thus, one of the major challenges facing designers today concerns the best way to utilize technology to serve the needs of society without exceeding the limits of those individuals who must operate the technology.

**What is Automation?** Automation has been described as a machine agent that can execute functions normally carried out by humans (Parasuraman & Riley, 1997). These can be entire functions, activities, or subsets thereof. Automation serves several purposes (Wickens, 1992). It can perform functions that are beyond the ability of humans, it can perform functions for which humans are ill-suited, and it can perform those functions that humans find bothersome or a nuisance.

**Levels of Human-Automation Interaction.** The level of automation in a system can vary. Sheridan and Verplank (1978) proposed a model where differences range from completely manual to fully automatic (see Table 1). Several examples of degrees of automation can be found in a typical automobile. At the lowest level, virtually all automobiles require the driver to put the car into gear. At the other extreme, the antilock braking system calculates how much pressure to apply to each wheel to bring the car to a halt without locking up any wheels. It does so without communicating any of its calculations or actions. All the driver has to do is apply the brakes. The presets on the car's audio system allow individuals to automatically tune to their favorite stations. The system limits the range of available frequencies to a select few and presents these choices to the user on separate buttons.



Table 1. 10 Levels of Human-Automation Interaction (Sheridan & Verplank, 1978)

1) Whole task done by human except for actual machine operation	Manual
2) . . .	Semiautomatic
3) . . .	Semiautomatic
4) Computer suggests options and proposes one of them	Semiautomatic
5) Computer chooses an action and performs it if human approves	Semiautomatic
6) Computer chooses an action and performs unless human disapproves	Semiautomatic
7) . . .	Semiautomatic
8) . . .	Semiautomatic
9) . . .	Semiautomatic
10) Computer does everything autonomously	Automatic

**Information Processing Framework.** Recently, Parasuraman, Sheridan, and Wickens (2000) expanded upon this model to provide designers with a framework for considering what types and levels of automation ought to be implemented in a given system. This expanded model allows for various levels of automation within different functions. The four functions they describe are system analogs of different stages of human information processing: information acquisition, information analysis, decision selection, and action implementation (see Table 2).

Table 2. Information processing functions (based on Parasuraman, Sheridan, & Wickens, 2000)

Stage of Processing	Functions
Information Acquisition	Detecting and registering input data
Information Analysis	Applying cognitive functions to the information (e.g., analyzing and summarizing, making predictions, inferences, modifying and augmenting information displays, etc.)
Decision Selection	Augmenting or replacing human selection of decision options
Action Implementation	Executing functions or choices of actions

## Impact of Automation Technology

**Advantages of Automation.** Wiener (1980; 1989) noted a number of advantages to automating human-machine systems. These include increased capacity and productivity, reduction of small errors, reduction of manual workload and fatigue, relief from routine operations, more precise handling of routine operations, and economical use of machines. In an aviation context, for example, Wiener and Curry (1980) listed eight reasons for the increase in flight-deck automation: Increase in available technology, such as the Flight Management System (FMS), Ground Proximity Warning System (GPWS), Traffic Alert and Collision Avoidance System (TCAS); concern for safety; economy, maintenance, and reliability; decrease in workload for two-pilot transport aircraft certification; flight maneuvers and navigation precision; display flexibility; economy of cockpit space; and special requirements for military missions.

**Disadvantages of Automation.** Automation also has a number of disadvantages. Automation increases the burdens and complexities for those responsible for operating, troubleshooting, and managing systems. Woods (1996) stated that automation is "...a wrapped package -- a package that consists of many different dimensions bundled together as a hardware/software system. When new automated systems are introduced into a field of practice, change is precipitated along multiple dimensions" (p.4). Some of these changes include: (a) adding to or changing the task, such as device setup and initialization, configuration control, and operating sequences; (b) changing cognitive demands, such as decreased situational awareness; (c) changing the role that people in the system have, often relegating people to supervisory controllers; (d) increasing coupling and integration among parts of a system often resulting in data overload and "transparency" (Billings, 1997); and (e) increasing complacency by those who use the technology. These changes can result in lower job satisfaction (automation seen as dehumanizing), lowered vigilance, fault-intolerant systems, silent failures, an increase in cognitive workload, automation-induced failures, over-reliance, increased boredom, decreased trust, manual skill erosion, false alarms, and a decrease in mode awareness (Wiener, 1989).

## **The Air Transportation System: Problems of Automation**

**The current outlook.** Civil aviation has reached and maintained a remarkable level of safety with less than one accident per million departures. Although laudable, the considerable growth anticipated in the National Airspace System (NAS) will appreciably tax the ability of the system to effectively handle the estimated doubling of air traffic to 50 million flights per year by 2010. The mathematics in human and economic cost outcomes are staggering. The National Transportation Safety Board (NTSB) released the U.S. aviation accident rates for 2000 that, although an overall 3.8 percent decrease in total accidents, there was a 7.3 percent increase in passenger fatalities (748 in 2000). The projections for out years in aviation accident statistics would be the equivalent of a commercial airliner crashing once every two weeks or 25 accidents per year with 1,000 fatalities. Furthermore, these considerations can be extended to the general aviation community whose fatal accident statistics parallel those of commercial aviation (592 in 2000). Taken together, while aviation is among the safest industries in the world, the increased traffic volume will assuredly lead to more accidents unless significant remediations are put into practice.

Estimates of 60 to 80% of accidents being due to human error suggest that a promising avenue to pursue would be an investment into the understanding of, and better support of, human-machine interaction; this is especially the case for human-automation interaction. Significant areas that work need be directed towards include improved system design, selection, and training of pilots and air traffic controllers (ATC) with automation. Foremost among them is a systematic uncovering of what the etiologies of human error are and the countermeasures that can be applied to address the telling and often deleterious impact that automation in the aviation domain has played in contributing to the increasing number of accidents and incidents.

**Effects of automation on aviation.** The effect that automation has had on aviation can be summed up as both a blessing and a curse. Clearly, automation has significantly improved the capability and efficiency at which humans fly to and from destinations throughout the world. The impressive aviation accidents statistics are certainly governed and attributable to the increase in automation. However, although airplanes fly more precisely and the NAS allows for more capacity, authorities caution that the time of ever increasing automation may have reached an apex. The literature and aviation databases, such as the NASA Aviation Safety Reporting System (ASRS), abound with research and anecdotal evidence substantiating this fact. The consensus in the human factors community, which is beginning to be shared with others within aviation, is that automation has not so much made the job of flying or managing aircraft easier; instead, it has transformed the role of the operator from active to passive monitoring of automated systems. The net effect of this has been a migration of human

performance issues from perceptual / motor to cognitive in nature. Wiener and Curry (1980) provided the cries of caution as to how automation was bringing with it a new set of problems, unforeseen in aviation, and that ushered in a new age of study within human factors. Today, the concerns expressed over 20 years hence are as relevant today as they were back in the 1980s. Indeed, perhaps because of the better picture we are painting of the human-automation canvas, now more than ever is research needed to address these issues and discover solutions to the problems that are the bane of modern aviation.

**Future outlook and new problems.** NASA (National Aeronautics and Space Administration), the FAA (Federal Aviation Administration), and Eurocontrol have begun programs of research and development with the objective of providing air traffic controllers, flightcrews, and air traffic managers with automation aids so as to increase terminal and enroute throughput and capacity. The trend is to allow flightcrews to determine user-preferred trajectory and in-trail separation within air traffic service provider (ATSPs) defined corridors of air traffic flow, also known as “free flight”. To become a reality, significant research will be needed to determine the system impact on human performance and safety as, more than ever before, automation will be key to the successful implementation of such a system. Of course, the human-automation interaction issues that currently exist will not magically go away, and the evolution of the air transportation system will profoundly challenge the cognitive sciences including human factors to ensure “human-centered” design.

**The Need for Research.** Therefore, there exists the impetus to address the issues that plague current automated systems. Amazingly, while many of these issues are known and acknowledged by those outside human factors, the economic pressures and competition makes progress slow and implementation a significant challenge. To eventually be successful, however, will require a watershed in thinking about human error and automation. There are many who still consider automation the panacea of, and not the contributor to, aviation safety. They see human error is the elemental concomitant of, rather than the symptom of, mismatches between the human, machine, and environment. That more automation, less human interaction equates to less error and more safety. However, Reason (1990) noted the logical flaws of this pointing out that we cannot objectively measure how many accidents were not caused because the human was “in-the-loop”. Instead, the human, as pilot or air traffic controller, represents the last “error filter” that can catch those latent and active “pathogens” that will eventually migrate down the system; this is the so-called “swiss cheese phenomenon”. The perspective of human, as error filter, rather than as error source, has many implications for the future of automation design. Foremost among them, the concept of “human-centered” automation design that is directed towards the support of humans (Billings, 1997) and understands that knowledge-based cognitive processing (Reason, 1990) of humans cannot be modeled or emulated by computer logic.

The case has been made and, although the jury is still out, humans will continue to interface with the increasing introduction of technology in the form of automation. The recognition of what makes a human element necessary also points to man’s limitations; that is, the inability of pilots and air traffic controllers to effectively monitor automated tasks (Mosier, Skitka, & Korte, 1994). Many have pointed out that the information processing capabilities of pilots and air traffic controllers to see patterns and to generate solutions (knowledge-based processing) makes them susceptible to automatic and controlled errors when engaged in skill- and rule-based processing. Mackworth (1958) noted that humans are not good at vigilant-type behaviors, such as observing automated system states or watching a radar scope. These types of behaviors lead to what O’kiris and Endsley called the “out-of-the-loop” performance problem that can lead to decreased vigilance, increased mental workload, lower situation awareness, and complacency. Complacency, in fact, is a natural result since it represents a “strategy” or behavior directly attributable to the intrinsic nature of automation. Automation tends to “work as advertised” (Woods, 1996) and, therefore, operators develop understandable trust in the automation. The numerous incidents and accidents in the ASRS (Pope & Bogart, 1992) earmark complacency as a contributing factor in accidents, such as Eastern Airlines L-1011 (1972) CFIT, Miami; China Airlines B-747 (1985), uncontrolled vertical descent from 41,000 to 9,500 ft. over Pacific Ocean; and American Airlines B-757 (1995) CFIT, Cali, Columbia. These accidents were in major ways caused by over-reliance on

automation. Parasuraman and Riley (1998) noted that of the “use, misuse, disuse, and abuse” of automation, automation-induced complacency poses the most significant threat to poor human-automation interaction.

## **Recent Physiological Factors Research**

**Research Taxonomy.** Research has been conducted that has focused on base research issues surrounding human-automation interaction in the aviation domain. The focus of the research can be broken down into five related, but distinct research areas: hazardous states of awareness, individual difference variates, training countermeasures, adaptive automation, and spin-off technologies.

## **Hazardous States of Awareness**

An analysis of incidents in the Aviation Safety Reporting System (ASRS) database reveals that civil transport flight crews often relate their mistakes to their experiences of certain states of awareness. Narratives in the incident database contain descriptions of crewmembers becoming “complacent” and succumbing to “boredom.” The reports further describe experiences of diminished attention, compromised vigilance, and lapsing attention, frequently not associated with fatigue. These experiences are variously attributed to conditions of quietness, droning noise and motion, monotony, repetition, and familiarity. Crewmembers have recalled that they were excessively absorbed or dangerously preoccupied prior to an error incident. These crewmembers, whose responsibility it is to monitor and manage the progress of highly complex and automated systems, occasionally lapse into awareness states that are incompatible with the demands of the task. These states can be characterized as being “hazardous.” As Billings (1991) notes, “Few tasks are more soporific than watching a highly automated vehicle drone on for many hours, directed by three inertial navigation systems all of which agree within a fraction of a mile.” Hazardous states of awareness occur most frequently under just such conditions.

Numerous studies and critical incident investigations have shown that hazardous states of awareness (HSAs) can lead to catastrophic consequences. Young & Hashemi (1991), for example, showed that truck driver fatigue may play a role in up to as many as 41% of accidents. As another example, loss of situation awareness have been found to be a major contributor to Controlled Flight Into Terrain (CFIT) accidents (Shappell & Wiegmann, 1997) and HSAs (i.e., problems of attention) may account for up to 40% of all CFIT accidents.

Attention may be usefully characterized by three aspects: (1) distribution (diffused versus concentrated); (2) intensiveness (alert versus inattentive); and (3) selectivity (the “what” of attention). Distribution and intensiveness are influenced by the state of awareness being experienced; the experiences cited above vary along these dimensions. Selectivity refers to the contents of awareness. In designing for effective integration of human and system, it is important to provide ready access to useful information so that the contents of awareness support informed action. It is also important to design for human involvement in system function to promote effective state of awareness (i.e., to promote consistent mental engagement in the supervisory task). Other ways of describing attention can be delineated into three major components of selection, vigilance, and control. But, at present, there is no major taxonomy of attention put forward to express the diversity of attention processes. However, a number of useful psychological constructs have emerged that have proved invaluable in guiding the research in the area.

**What are HSAs?** Hazardous states of awareness can be differentiated into six related constructs: blocks, task-unrelated thoughts, lapses and slips, complacency, mental fatigue, and boredom. Based on the ASRS narratives, pilots make frequent reports of “daydreaming”, “worrying”, “distracting”, “hypnotized”, “losing focus”, “totally absorbed our [sic] concentration and forgot the big picture”, “tunneling”, “channelization”, “complacent”, and “preoccupation” to name a few of their words. These

descriptions have been captured and subsumed by the development of a HSA taxonomy that use these six psychological constructs.

**Blocks.** The psychological construct of a block is similar to the layperson's attribution to the word. We often talk of "having a mental block" that doesn't allow us to accomplish a said task. The term was first used by Bills (1931) to refer to subjects that occasionally take a very long time to name certain colors. During a block, information is available about a person's state of awareness. However, it is contradictory in that a person doesn't cease responding but returns to a task after a period of time suggesting a "preoccupation" or a redirection of attention elsewhere. Therefore, a block is likened to a task-unrelated thought (TUT) and, therefore, blocks are related to arousal and attentive processing. But, there is also evidence that the frequency of blocks increases over time and are more likely with repetitive, boring tasks such as vigilance or supervisory control tasks. Therefore, blocks are related to low arousal or boredom although there is evidence that boredom is actually highly arousing and high in workload (Scerbo, Greenwald, and Sawin, 1995). Makeig & Inlow (1993) have noted that with blocks, while different than microsleep episodes, there is increased delta activity in the electroencephalogram (EEG) record. Therefore, blocks are distinct from lapsing into daydreaming or a sleep-like state. Rather, studies have shown that vigilance or sustained attention is most at work in the onset of blocks. For example, Robertsen et al. (1997) found that on several tests of "everyday" attention, sustained attention tests correlated significantly with the frequency of blocks. Furthermore, Smith and Nutt (1996) reported that clonidine, that reduces the noradrenaline release at the presynaptic level, can increase the incidence of blocks whereas idazoxan, a noradrenaline agonist, reverses the effect.

**Task-Unrelated Thoughts.** If subjects are not asleep during a block, they may be engaged in thoughts not directly related to the task at hand. Giambra (1993) characterized the state as consisting of *task-unrelated thoughts* or (TUTs). These TUTs can be conceived as a form of daydreaming, (i.e., thoughts that flirt from topic to topic). However, recent evidence has shown that it is equally plausible that "preoccupation," or focused attention on a single thought, might also qualify as a TUT. Antrobus, Coleman, and Singer (1967) first reported on the role of TUTs in under long-duration task performance, and they administered a daydreaming questionnaire to a sample of subjects and divided them into high and low daydreaming groups. Those who scored high in daydreaming showed a greater vigilance decrement over time and they also reported more irrelevant thoughts and indicated that the incidence of these increased with time on task.

Giambra (1993) argued that TUTs were a specific subject state reflecting internal dispositions and, therefore, would qualify as a hazardous state of awareness. The aspect of attention that is probably most closely involved in TUTs is vigilance although selection is presumably also involved to an extent. It has been postulated that a TUT may reflect a failure of the selection mechanism although it is possible that such occasional and random deviations may be a general characteristic of a selective attention mechanism, as originally proposed by Broadbent (1958) in his well-known filter theory. As an example, in a dichotic listening study in which subjects are asked to attend to stimuli in one ear, participants report that they are sometimes aware of stimuli from the other ear. If this reflects random but infrequent sampling of irrelevant channels by the attention filter, TUTs may just be a natural consequence, like noise in a sensory channel. Therefore, it is possible that TUTs may not represent active failures of selection but instead represent the normal, noisy operation of selective attention.

**Lapses and Slips.** Task-irrelevant thoughts can have an effect on a person's goal-directed behavior, but may not directly lead to actions themselves or have other direct behavioral consequences. Such thoughts are likely to be prevalent whenever a person is engaged in a repetitive, habitual activity for sustained periods of time under relatively low external task demands. Examples of this would include performing routinized tasks during cruise phase of flight. Unlike TUTs, however, performance of these tasks might also result in erroneous actions because the familiar and routine nature of the task promotes what has been termed "automatic processing." When these actions are unintended, they have been referred to as either lapses or slips (Norman, 1981; Reason, 1984).

Norman (1981) defined slips as those actions that do not agree with a person's intentions. Reason (1984) defined slips in much the same way and further defined lapses as intention failures that are not necessarily apparent in behavior. Lapses and slips are both seen as an outcome of a failure of an attentional check (monitoring) mechanism when a person is engaged in a habitual or routine activity. Reason (1990), illustrate the concepts of slips and lapses:

Slip: "I had decided to cut down on my sugar consumption and wanted to have my cornflakes without it. **But the** next morning, however, I sprinkled sugar on my cereal, just as I always do."

Lapse: : "I walked to my bookcase to find the dictionary. In the process of taking it off the shelf, other books fell onto the floor. I put them back and returned to my desk without the dictionary. "

Reason (1990) argued that slips and lapses represent a fundamental failure of attentive monitoring. He proposed that skilled actions carried out in a familiar environment result in automatic execution of stimulus-response (SR) sequences with only an occasional attentional check.

**Complacency.** If we can operationally define a slip as representing a failure of attentive monitoring of one's own action sequences during performance of habitual tasks, *complacency* represents a failure to monitor the actions of a machine or computer that may also be under habitual and familiar circumstances. The term complacency has been used for some time in aviation and it has been inferred when the pilot was thought to be irresponsible in not checking on some aircraft subsystem or in not adequately monitoring the cockpit instruments (Hurst & Hurst, 1982). Complacency has become the focus of even more recent research because of increased automated systems in the cockpit (Billings, 1996). However, the term has been applied in many other domains where automation is being increasingly implemented, such as shipping (National Transportation Safety Board, 1997), process control (Lee, Parasuraman, & Bloomfield, 1997), and driving (DeWaard & Brookhuis, 1999). Automation-related complacency has been defined to refer to a condition of excessive trust in an automated subsystem; that trust doesn't allow them adequately monitor the system for anomalies or malfunctions (Lee et al., 1997; Parasuraman, Molloy, & Singh, 1993; Singh, Molloy, & Parasuraman, 1993; Wickens, Mavor, & McGee, 1997).

Wiener (1981) defined *complacency* as "a psychological state characterized by a low index of suspicion." Billings, Lauber, Funkhouser, Lyman, and Huff (1976), in the Aviation Safety Reporting System (ASRS) coding manual, defined it as "self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state." The condition is surmised to result when working in highly reliable automated environments in which the operator serves as a supervisory controller monitoring system states for the occasional automation failure. It is exhibited as a false sense of security, which the operator develops while working with highly reliable automation; however, no machine is perfect and can fail without warning. Studies and ASRS reports have shown that automation-induced complacency can have negative performance effects on an operator's monitoring of automated systems (Parasuraman, Molloy, & Singh, 1993).

Although researchers agree that complacency continues to be a serious problem, little consensus exists as to what complacency is and the best methods for measuring it. Nevertheless, after considering the frequency with which the term "complacency" is encountered in the ASRS and analyses of aviation accidents, Wiener (1981) proposed that research begin on the construct of complacency so that effective countermeasures could be developed.

One of the first empirical studies on complacency was Thackray and Touchstone (1989) who asked participants to perform a simulated ATC task either with or without the help of an automated aid. The aid provided advisory messages to help resolve potential aircraft-to-aircraft collisions. The automation failed twice per session, once early and another time late during the 2-hr experimental session. These researchers reasoned that complacency should be evident and, therefore, participants would fail to

detect the failures of the ATC task due to the highly reliable nature of the automated aid. However, although participants were slower to respond to the initial failure, reaction times were faster to the second automated failure.

Parasuraman, Singh and Molloy (1993) reasoned that participants in the Thackray and Touchstone (1989) experiment did not experience complacency because of the relatively short experimental session and because the participants performed a single monitoring task. ASRS reports involving complacency have revealed that it is most likely to develop under conditions in which the pilot is responsible for performing many functions, not just monitoring the automation involved. Parasuraman et al. (1993) suggested that in multi-task environments, such as an airplane cockpit, characteristics of the automated systems, such as reliability and consistency, dictate how well the pilot is capable of detecting and responding to automation failures. Langer (1989) developed the concept of premature cognitive commitment to help clarify the etiology of automation-induced complacency. According to Langer,

When we accept an impression or piece of information at face value, with no reason to think critically about it, perhaps because it is irrelevant, that impression settles unobtrusively into our minds until a similar signal from the outside world – such as a sight or sound – calls it up again. At that next time it may no longer be irrelevant, most of us don't reconsider what we mindlessly accepted earlier.

Premature cognitive commitment develops when a person initially encounters a stimulus, device, or event in a particular context; this attitude or perception is then reinforced when the stimulus is re-encountered in the same way. Langer (1989) identified a number of antecedent conditions that produce this attitude, including routine, repetition, and extremes of workload; these are all conditions present in today's automated cockpit. Therefore, automation that is consistent and reliable is more likely to produce conditions in multi-task environments that are susceptible to fostering complacency, compared to automation of variable reliability.

Parasuraman, Singh and Molloy (1993) examined the effects of variations in reliability and consistency on user monitoring of automation failures. Participants were asked to perform a manual tracking, fuel management, and system-monitoring task for four 30-minute sessions. The automation reliability of the system-monitoring task was defined as the percentage of automation failures that were corrected by the automated system. Participants were randomly assigned to one of three automation reliability groups, which included: constant at a low (56.25%) or high (87.5%) level or a variable condition in which the reliability alternated between high and low every ten minutes during the experimental session. Participants exhibited significantly poorer performance using the system-monitoring task under the constant-reliability conditions than under the variable-reliability condition. There were no significant differences between the detection rate of the participants who initially monitored under high reliability versus those who initially monitored under low reliability. Furthermore, evidence of automation-induced complacency was witnessed after only 20 minutes of performing the tasks. Parasuraman et al. (1993) therefore concluded that the consistency of performance of the automation was the major influencing factor in the onset of complacency regardless of the level of automation reliability.

Singh, Molloy, and Parasuraman (1997) replicated these results in a similar experiment, which examined whether having an automated task centrally located would improve monitoring performance during a flight-simulation task. The automation reliability for the system-monitoring task was constant at 87.5% for half the participants and variable (alternating between 56.25% and 87.5%) for the other half. The high and low constant groups were not used in this study because participants in previous studies were found to perform equally poorly in both constant reliability conditions. A constant high level of reliability was used instead because complacency is believed to most likely occur when an operator is supervising automation that he or she perceives to be highly reliable (Parasuraman et al., 1993). Singh and his colleagues found the monitoring of automation failure to be inefficient when reliability of the

automation was constant but not when it was variable, and that these failures cannot be prevented by locating the task in the center of the computer screen. These results indicate that the automation-induced complacency effect discovered by Parasuraman et al., is a relatively robust phenomenon, which is applicable to a wide variety of automation reliability schedules.

**Mental Fatigue.** Each of the four HSAs discussed so far have one common theme to them and that occur whenever a person is engaged in habitual tasks for considerable amounts of time. Any condition such as that can also produce a state of *mental fatigue*. There are a number of different definitions of what mental fatigue really is, but doubtless its effects on aviation safety are well known and appreciated. Bartlett (1943) proposed that fatigue represented measurable changes in performance arising from extended what he termed, "time on task." However, Bartley and Chute (1947) noted that actually fatigue is a psychological (as opposed to physiological) consequence of prolonged activity and requires subjective appraisals of feelings of tiredness as well. The subjective dimension is currently viewed as being the more significant factor in the onset of mental fatigue. Brown (1993), for example, defined it as a "disinclination to continue performing a task because of perceived reductions in efficiency."

What is the role of attention in fatigue? One intuitive notion that has long held currency is that mental fatigue is associated with a depletion of attentional resources, or an increasing inability to allocate attentional resources to task performance. Resource theory (Wickens, 1984) posits that a person has a limited supply of attention capacity available and that performance decrement occurs when task demands on capacity exceed the supply available. The nature of the decrement is dependent on the strategies the person uses to allocate attentional resources to different tasks or to components of a task. Fatigue may impair these processes.

**Boredom.** To understand mental fatigue, it requires also an understanding of boredom. Many have defined boredom in several ways, but almost all include the words "state", "feeling", and/or "conflict" in them. Fenichel (1951), for example, described boredom as a feeling of displeasure caused by a conflict between a need for intensive psychological activity and an unstimulating environment. Whereas Barmack (1939) described boredom as a feeling that produced a sleep-like state resulting from a conflict between continuing and removing oneself from an unpleasant situation. Others have described boredom as a physiological state associated with arousal (Berlyne, 1960; Hebb, 1955) or an emotion (Damrad-Frye & Laird, 1989).

One mechanism for boredom may be habituation in which receptivity to an event declines with exposure to a steady, prolonged stimulus (Kroemer & Grandjean, 1997). The postulated benefit is that it helps protect the central nervous system (CNS) from becoming overloaded or saturated with impulses from the peripheral sense organs (e.g., eyes). Without habituation, a person would be in a state of heightened arousal all the time and, therefore, in a monotonous situation, habituation protects the individual from becoming inundated with repetitive stimuli. However, a consequence is that it produces boredom. O'Hanlon (1981) suggested that habituation may be the psychophysiological beginning of boredom. It may lower the arousal level of an individual when presented with repetitive stimuli (Lynn, 1966) and when an individual first encounters a task, the stimulation produced by the situation will produce arousal at or above optimal levels. O'Hanlon argued that if the arousal level were to fall below the minimum level required to achieve par performance, an effect would be an increase in lapses, mental blocks, errors, and detection failures in monitoring tasks. In order to attempt to prevent habituation and maintain attention, considerable effort would have to be expended and this is seen as unpleasant. Therefore, boredom may be viewed as the conflict between habituation and sustained effort to maintain a sufficient level of arousal to perform adequately at a task (O'Hanlon, 1981).

## Research on Hazardous States of Awareness

Research has been conducted to examine the etiologies of hazardous states of awareness. The focus has been on developing the necessary tools and insights for classification of various HSAs, so that



countermeasures can be developed. The term, “hazardous states of awareness” has been endorsed by the aviation community as a catch-all phrase for the phenomenological states that pilots may enter that can prove deleterious to aviation safety. For example, the FAA (1996) in their national plan called on more research to be conducted to understand the nature of these hazardous states of awareness so that more effective training and system design can be developed to help limit the estimated 60-80% of accidents caused by human errors.

There exists a considerable knowledge base of basic research from various research communities that we can draw upon in conducting human factors, applied research that has been the primary focus of recent research at the NASA Langley Research Center. Nevertheless, in addition to the content analyses conducted on the ASRS database, there has also been performed a considerable amount of basic research into the causes of HSAs.

**Task-Unrelated Thoughts.** Research was conducted at NASA Langley Research Center to examine EEG and Event-Related Potential (ERP) correlates of different levels of attention that were induced by different task demands (Cunningham, S. C. & Freeman, F. G., 1994). The aim of the study was to further investigate the relationship between fluctuations in vigilance performance and electrocortical activity. Our hypotheses were that the level of difficulty of a task would differentially consume attentional resources (Wickens, 1994) and, in turn, significantly impair operator vigilance performance (e.g., ability to monitor). Participants performed a vigilance-type task for 40-min while their EEG and performance data was gathered, and the task required responses to events (30 or 60 event rate) with 1 critical event per minute. The results evinced a significant N1-P2 relationship to task-irrelevant tones (Figure 1; Makeig & Inlow, 1992), reaction time (RT), and A' significantly discriminated between task difficulty (see Table 1). There was also a significant increase in Task-Unrelated Thoughts (TUTs) as time-on-task increased. An analysis of the absolute EEG indices for pre- and post-TUTs did not yield any significant differences. However, EEG indices of arousal (e.g., beta/(alpha+theta)) demonstrated significant differences for all parietal lobe sites (Pz, P3, and P4). The findings of the study suggest that daydreaming and/or TUTs is reflected in the EEG in terms of ratios of power in beta, alpha, and theta bandwidths.

Table 1: Mean Performance Across Periods

Proportion of Hits	0.94 (.08)	0.84 (.01)	0.77 (.03)	0.76(.06)
Response Time to Hits	836.8 (141)	955.6 (164)	990.9 (170)	990.9 (170)
Perceptual Sensitivity (A')	0.951 (.03)	0.923 (.04)	0.916 (.04)	0.909 (.05)
Response Criterion (B'')	-0.43 (.78)	0.22 (.76)	0.36 (.69)	0.18 (.85)
Number of TUTs	6.5 (5.6)	10.4 (11.3)	10.5 (10.7)	10.9 (11.8)

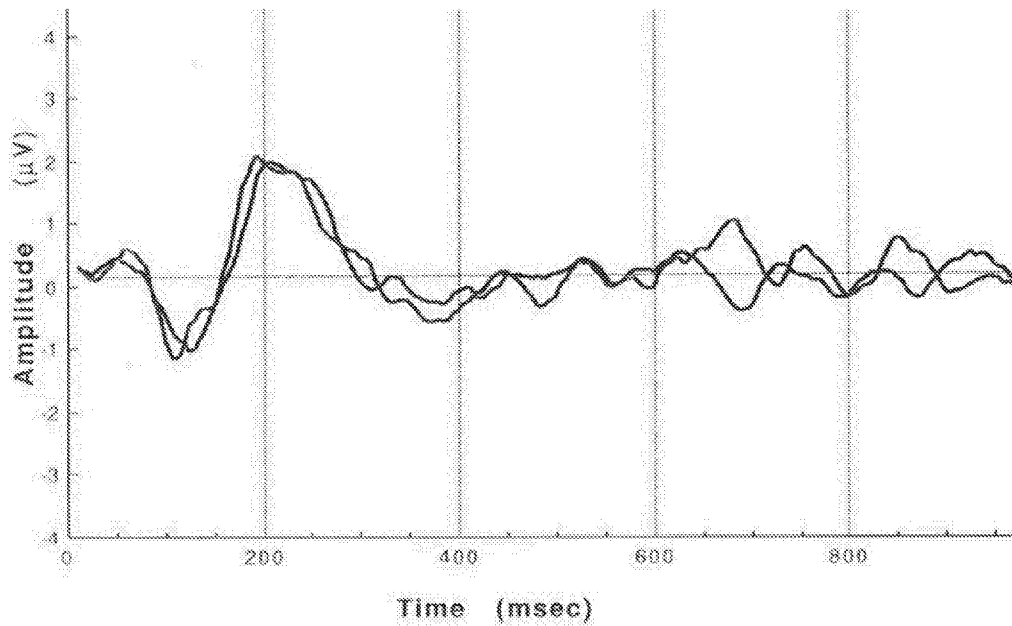


Figure 1. ERP waveform for Irrelevant, Odd-Ball Tones (Pz & Cz)

**Neural Network EEG Mental Workload Classification.** Research has been conducted to investigate the ability of the neural network to classify levels of cognitive workload using EEG band activity and neural networks. Five subjects (three males and two females) were asked to perform tasks from the Multi-Attribute Task Battery (MATb). A Physiological Factors team member, Dr. Raymond Comstock of Langley Research Center, developed the MATb which provides a benchmark set of tasks for use in a wide range of laboratory studies of operator performance and workload (Arnegard & Comstock, 1991; Figure 2). Each subject experienced two combinations of workload (either low, medium, or high) with baseline data collected prior and subsequently to the six experimental trials. Six channels of EEG data were recorded at a sampling rate of 100 Hz at electrode sites: C3, Cz, C4, P3, Pz, and P4. All eyeblinks and other artifacts (e.g., EMG) were removed prior to data analysis.

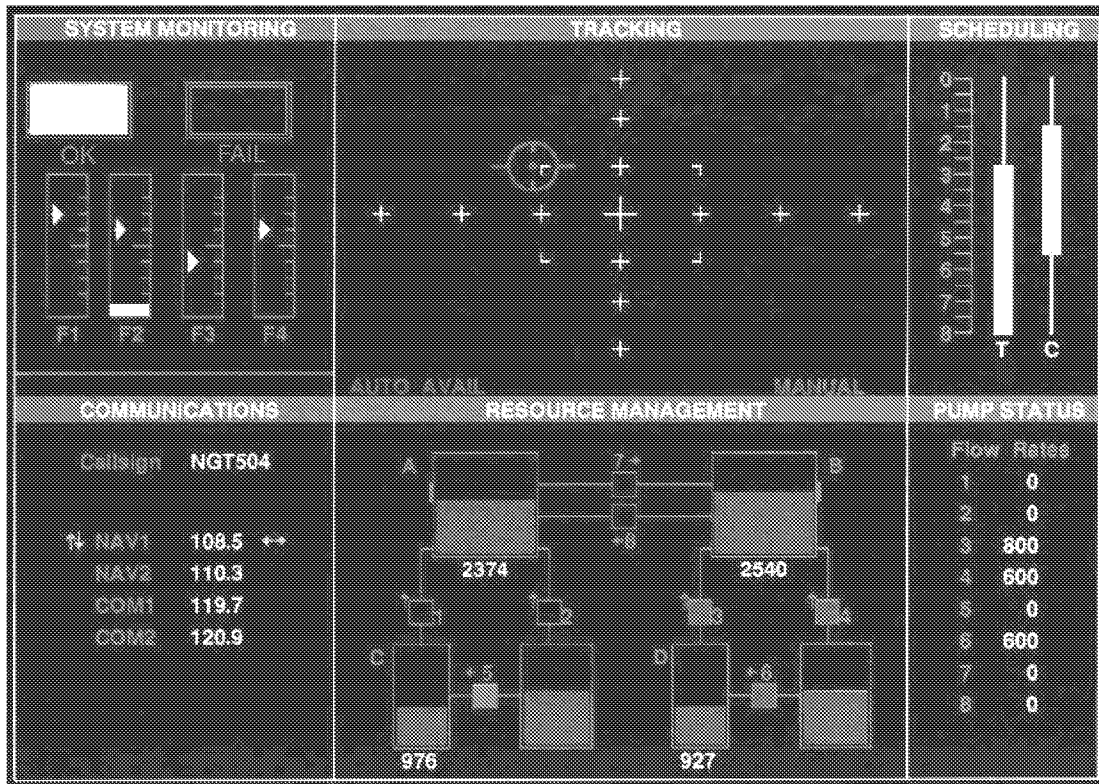


Figure 2. The Multi-Attribute Task Battery (MATb)

In ten-second segments, each signal was bandpass filtered using a bank of ideal filters to produce five bands of EEG: delta (DC-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30Hz), and ultra-beta (31-42Hz). Data was overlapped using an 80% overlap moving window. Log power of the bands were calculated which resulted in 30 features to be used as inputs to the neural network. Four minutes of data were used in each trial workload resulting in a total of 462 samples per person.

A resilient feed-forward neural network was used in this study. Adaptive learning and momentum were used to decrease the time required for training the network. The network architecture, shown in Figure 3, consisted of three layers. The activation function for all layers were logistic sigmoid functions in the form  $f(a) = (1 + e^{-a})^{-1}$ . In order to compute the average classification accuracies, four runs were completed per subject.

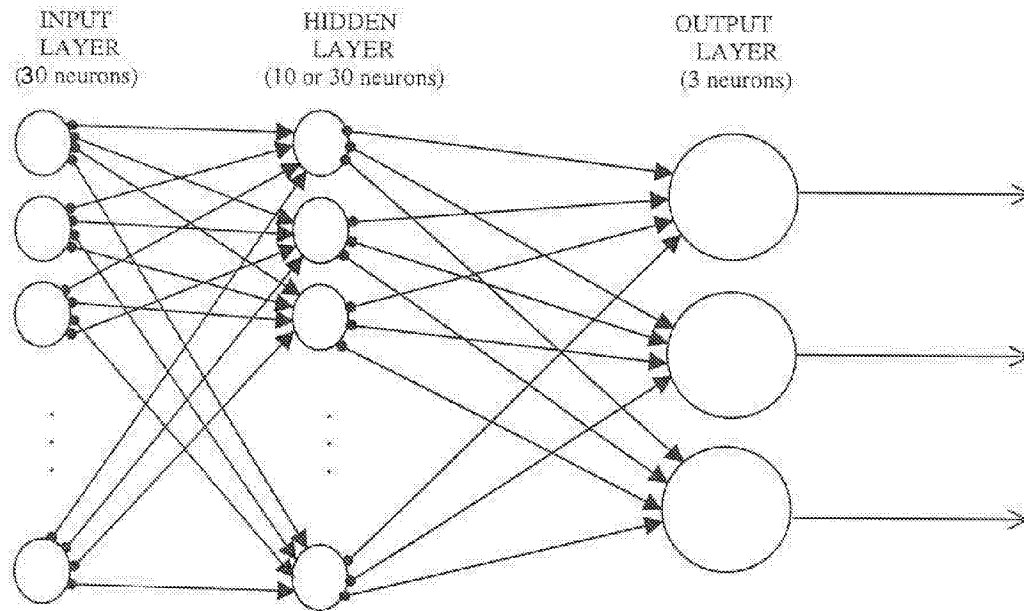


Figure 3. Neural Network Architecture

The data were segmented into two groups: training (80%) and testing (20%). The exemplars for each were chosen randomly per subject. The data was normalized between zero and one. Weights and biases were initialized using the Nguyen-Widrow random generator for logistic sigmoid neurons. The minimum validation error was the primary stop criterion and was set at a mean-square error of 0.02. This generally occurred in less than 60 epochs or passes through the data. Batch processing was used. Target vectors of length three were constructed as follows:

LOW [0.9 0.1 0.1] MEDIUM [0.1 0.9 0.1] and HIGH [0.1 0.1 0.9]

The results of the study showed that the workload conditions were correctly classified at rates above the 84% on the average. In addition, the classification accuracy across subjects varied less than 10% per subject. Accuracy levels increased only slightly from the first trial of a particular workload to the second trial of that same workload. Tables 2-4 and Figure 4 present these results. The general implication of these findings is that a resilient backpropagation neural network based on EEG can be used to classify high levels of workload and can be applied to the assessment of hazardous states of awareness. Current research has been focused on the further development of these neural network algorithms and the use of a wavelet transformation method using multi-resolution analysis of the time-frequency signal of both the EEG and ECG to be used in adaptive automation system design (see below).

Tables 2-4. Subject X Workload, Subject X Trial, and Workload X Trial Accuracies

SUB/ WL	1	2	3	4	5
LOW	0.8832	0.8804	0.8967	0.8043	0.8370
MEDIUM	0.8723	0.8723	0.9402	0.9185	0.9130
HIGH	0.8641	0.8478	0.9457	0.8505	0.8370

SUB/ TRIAL	1	2	3	4	5
PRE-	0.8986	0.8986	0.8986	0.8696	0.8406
POST-	0.8623	0.8949	0.9529	0.8768	0.9275
COMBINED	0.8659	0.8370	0.9293	0.8424	0.8406

WL/ TRIAL	1	2	3
PRE-	0.8652	0.8978	0.8804
POST-	0.8826	0.9217	0.9043
COMBINED	0.8467	0.8967	0.8457

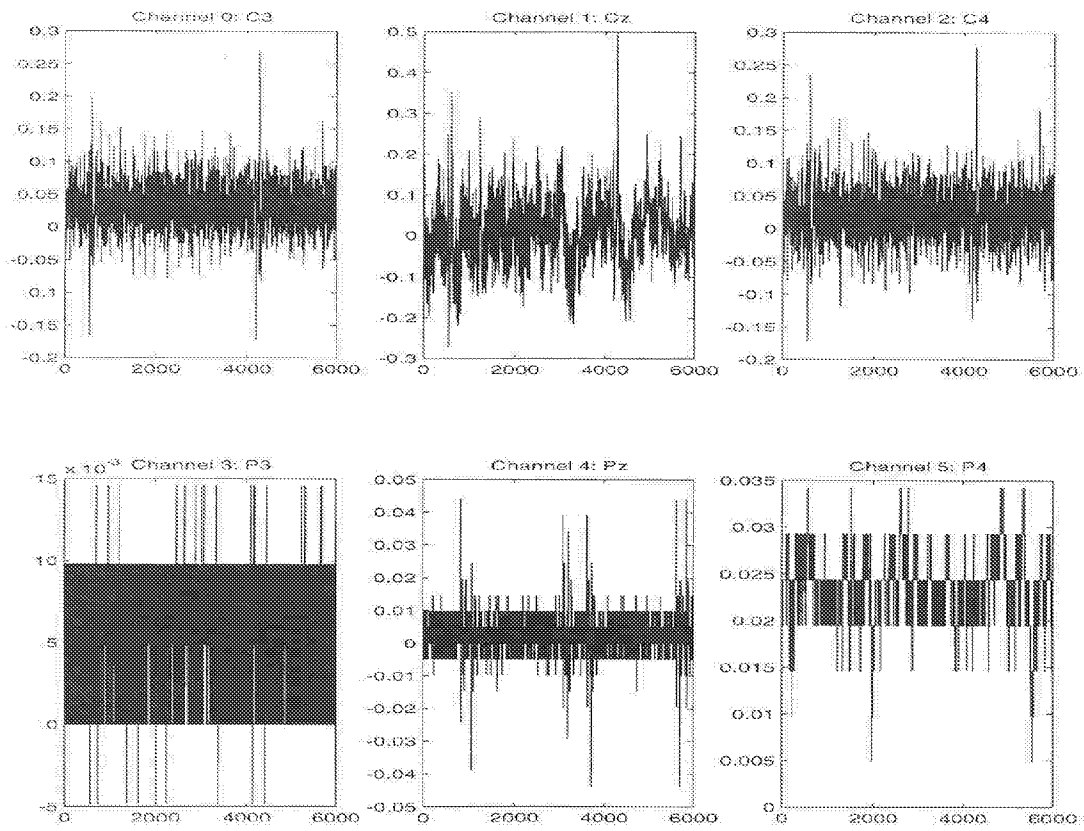


Figure 4. EEG Band Activity Across Electrode Sites

#### **Hazardous States of Awareness Assessment Using Finger Tremor and EEG Research.**

Healthy individuals exhibit a rhythmic physiological tremor that has been studied for at least 100 years. However, few studies have addressed the tremor/EEG relationship. Spectral analyses of these micro-tremors from the finger suggest components at approximately 8-15 Hz and 22-28 Hz. In this study, we sought to determine the relationship between tremor and electrocortical processes and well as the use of tremor to reflect processing demands. Dr. Bill Ray, of Penn State examined 9 healthy individuals during a baseline, a mental math and a relaxation task of 2-minute duration each. Tremor measurements were taken with an accelerometer attached to the right and the left index finger. EEG measurements were recorded from 15 cortical sites referenced to linked ears. Horizontal and vertical eye movement was also recorded. The tremor frequency found in the left and right hands was identical with a peak at approximately 10.5 Hz and 15.5 Hz. During the three tasks the frequency of tremor remained constant with the amplitude increasing between the mental math and relax task. There was a .2 coherence at 15.5 Hz between the two hands at baseline which decreased to .11 during relaxation and .02 during mental math suggesting that tremor may be influenced by different processes during different tasks. EEG frequency at the central sites showed a peak of 8.2 Hz during tasks resulting in low EEG/tremor coherence. As in previous research EEG hemispheric differences varied according to task. The work has been applied to the use of the EEG/tremor relationships index of cognitive/mental workload.

**EEG Density Measures of Mental Fatigue.** Research was conducted at NASA Ames Research Center that examined the psychological construct of mental fatigue and the use of a mathematical application to measure EEG energy density as a potential metric of mental fatigue. Mental fatigue often poses a serious risk, even when performance is not apparently degraded. When such fatigue is associated with sustained performance of a single type of cognitive task it may be related to the metabolic energy

required for sustained activation of cortical fields specialized for that task. The objective of this study was to adapt EEG to monitor cortical energy dissipation at a functionally specialized site over a long period of repetitive performance of a cognitive task.

Multi-electrode event related potentials (ERPs) were collected every fifteen minutes as 9 subjects performed a mental arithmetic task (algebraic sum of four randomly generated negative or positive digits). A new problem was presented on a computer screen 0.5 seconds after each response; some subjects endured for as long as three hours. The ERPs were transformed to show a quantitative measure of scalp electrical field energy. The average energy level at electrode p3 (near the left angular gyrus), 100-300 msec latency, was compared over series of ERPs.

The results showed that, for most subjects, scalp energy density at p3 gradually fell over the period of task performance and dramatically increased just before the subject was unable to continue the task. This neural response can be simulated for individual subjects using a differential equation model in which it is assumed that the mental arithmetic task requires a commitment of metabolic energy that would otherwise be used for brain activities that are temporarily neglected. Their cumulative neglect eventually requires a reallocation of energy away from the mental arithmetic task.

This research demonstrates a method for studying cognitive fatigue, independent of the subject's manifest performance. It also suggests that scalp energy density EEG data may reflect changes in cortical metabolic energy distributions, similar to PET and other scanning methods. There are two intriguing aspects of these results. First of all, they suggest that this approach to monitoring mental fatigue may have practical use. If one were to follow the rule that a task should be terminated as soon as an unusually strong peak of energy density appears at electrode site p3, that rule would have avoided operator failure in 7 of the 9 cases. The second intriguing aspect of the results is that they seem to fit a plausible hypothesis about brain function. The initial downward drift in energy density at p3, which characterized five of the subjects, is consistent with the assumption that other brain activities were making progressively greater demands upon available metabolic energy. These may have been homeostasis functions. Or they may have been distractions, not least of which would have been such things as muscle cramps and bladder distention. That the downward drift is oscillatory would be consistent with the possibility that some of the neglected brain functions competing for energy could be adequately serviced by a brief commitment of energy. A periodic 'clean-up' of this sort can be compared to such household chores as taking out the garbage or mowing the lawn.

There are two implications of this research. One is the value of energy density ERPs in psychophysiological research, especially in what might be called cognitive "ergodynamics." The other is the importance, for future research, of the hypothesis that mental fatigue can be studied as an aspect of metabolic energy resource allocation.

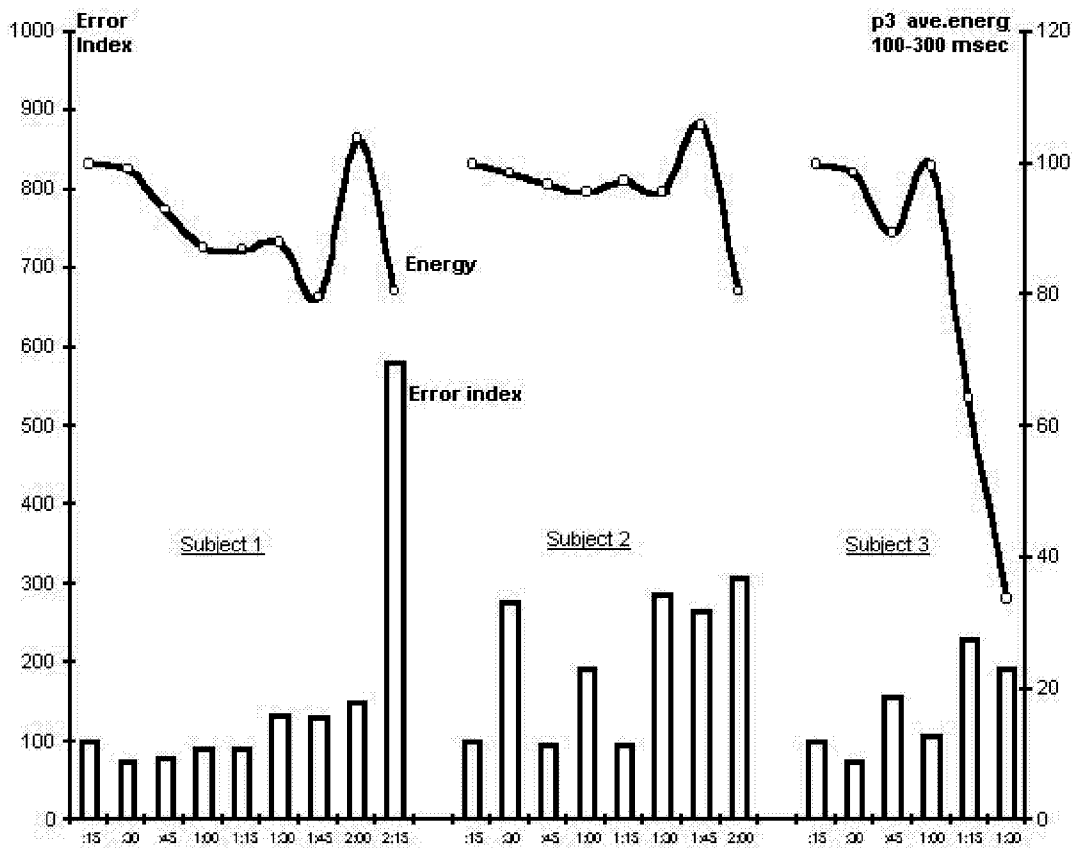


Figure 5. Three Subject's Showing Characteristic Energy Density Pattern of Mental Fatigue

## Implications of HSAs Research: Development of Countermeasures

NASA has been committed to conducting research into the etiologies of hazardous states of awareness that has been shown through numerous case studies to be a significant barrier to achieving NASA's goal of reducing the aircraft accident rate by a factor of 5 within 10 years and a factor of 10 within 20 years. Such a goal is going to require that considerable research be directed at understanding the underlying causes of why pilot make errors. Furthermore, it is not enough to just identify why pilots make the errors they do, but to also develop countermeasures that can "catch" them before they are made or, at the very least, minimize their consequences.

As an example, consider the HSAs of automation-induced complacency. Pilots remark that upwards of 50% aviation accidents are caused by complacency (Jensen, 1995). Patiky (1991) stated that, "When diligence and skepticism fade in the glow of self-satisfaction, a pilot is in for a rude awakening." Wiener and Curry (1982) highlighted the problem in discussing the ever increasing "technology creep" into modern aviation. As pilots increase in monitoring, supervisory control functions and decrease in manual control functions, these eminent researchers predicted that the prevalence of boredom, erosion of competence, and complacency will begin to appear in the literature. That very year, an Air Florida B-737 crashed at Washington (now Reagan) National Airport shortly after takeoff in snow conditions that killed 74 of 79 persons on board. A contributing factor in the accident was the overconfidence in automation and the reluctance of the first officer to note the engine pressure indicator (EPR) values as anomalous.



The next year, 1983, a B-747 was destroyed by air-to-air missile as it ventured into Russian airspace because the flightcrew inadvertently left the autopilot in heading mode rather than INS mode. In 1984, a DC-10 crashed at J.K. Kennedy Airport because of the flightcrews disregard for published procedures for monitoring and controlling airspeed during the final approach, overreliance on autothrottle speed control system (which has a history of malfunctions), and cognitive overhead associated with executing a missed approach. The National Transportation Safety Board (NTSB) noted that “performance was either aberrant or represents a tendency for the crew to be complacent and over-rely on automated systems” (NTSB, 1984). The NTSB discussed the problem of overreliance and complacency on automated systems at length in the report. Nevertheless, despite a growing awareness in the industry, accidents such as these have gone unabated and one can continue the timeline onwards to present day and see that every year shows signs that automation-induced complacency, as do other HSAs, continues to be a major issue for aviation safety. So, what can be done?

A sample of human factors experts and research conducted in the area of aerospace has identified that there are three focus areas that can have a dramatic impact on reducing the incidence of HSAs in the aviation domain: Selection and individual difference variates, training, and automation design. Below are descriptions of these focus areas, which then follow with detailed research conducted at NASA Langley Research Center and colleagues outside the center to address shortcomings within these areas.

**Automation Redesign.** The current approach to automation has been “technology-centered” rather than “human-centered” (Billings, 1997). Satchell (1998) has stated that the outcome of this has been that feedback has lagged behind autonomy and authority, comprehensibility has lagged behind dependability and flexibility, and creativity has lagged behind conformity and ritual. Human-centered automation design attempts to increase the “sharing of purpose by humans and machines” (Palmer et al., 1993) and seeks to mitigate the “peripheralization” that has come about as a result of technology-centered automation design. Peripheralization has come to be the “catch-all” for problems on the flightdeck including loss of situation awareness, complacency, miscommunication and intent inferencing, primary-secondary task inversion, automation deficits, boredom-panic syndrome, cocooning. Satchell (1998) notes that the current trend to “fix” the problem is unwise as a more fundamental rethinking is necessary of the human-automation interface.

A trend in this direction is a new form of automation and a reconceptualization of levels of automation and automation management sharing (Sheridan, Parasurman, & Wickens, 2000). Fitts (1951) put forward the task allocation “list” that dichotomized what machines and what humans are good at doing. Today, the thinking has changed considerably as numerous accidents on the flightdeck point that the traditional function allocation scheme no longer works in today’s supervisory control environment. The collective disquiet has grown louder and louder with more and more accidents being caused by poor human-automation mix. Therefore, “adaptive automation” has been touted as a remediation to these concerns and an attempted endorsement of an automation philosophy that is “human-centered”. There are a number of ways of looking at types of human-machine sharing and they range from direct, manual control to autonomous operation. However, these distinctions have been criticized as being too subtle and simplistic and do not retain the true human-machine tapestry of shared control (Tenney et al., 1995). Adaptive automation seeks to blur these distinctions even further. In adaptive automation, the level or mode of automation can be modified or changed in real-time to account for changes in levels of operator workload. The key to automation is that both the human and the machine share control over the state of automation and, therefore, the aircraft. The notion of shared control is not new as the A-320 fly-by-wire aircraft has shared control of all normal flight control modes. However, the “shared control” of the A-320 acts more as an “invisible hand” that enacts control laws or forcing functions if pilot input exceeds performance envelopes. Adaptive automation is different than this in that the “intelligence” of the automation is able to determine taskload levels and make modifications to the task allocation mix in order to keep the pilot “in-the-loop.” The net outcome is posited to be that pilots will remain engaged and attentive and won’t succumb to the hazardous states of awareness associated with traditional automation. Still, many (Scerbo, 1996, Woods, 1996) have voiced concerns over automation and the need to research

the myriad issues associated with the revolutionary approach to improving human-automation interaction across the elements of the airspace system.

**Individual Difference Variate Research.** One potential area is to better examine how personality characteristics can influence the tendency to trust and over-rely on automation. Selection has come along way since the Wright Brother decided who would be first to fly by flipping a coin. There are a number of selection batteries, test, and instruments that attempt to select those men and women best suited for the demands of aviation. Selection is important and cost-effective. Consider that USAF training can cost upwards of 1 million dollars and there is currently a 30% dropout rate (Siem, Carretta, & Mercatante, 1988). Further consider that the loss of a modern commercial transport aircraft can be over 100 million not including the human costs of losing an average of 200-400 passengers. Therefore, developing more sophisticated selection tests could make for a much better safety and economical bottom-line.

Rippon and Manuel (1981) described the pilot as a high-spirited person who "...seldom takes his work seriously but looks upon 'Hun-strafting' as a great game..." and returns after a day of flying to drink, dance, play music, and cards. The "fly by the seat of your pants" pilot was necessary and desirable during WWI and WWII. The modern pilot, however, requires a much different person that is a "team-player." The selection of pilots has utilized many different methods in order to select that "right kind of person." However, to date, it is still not know just what that is. Houston (1988) stated that personality measures may make the most significant contribution to the prediction of post-training performance. Hunter and Burke (1995), however, stated that "...clearly much additional research is needed in this area to clarify the role of personality and its measurement..." Although there are a number of useful personality instruments for aviator selection, many of them employ traditional measures of personality including the MMPI, State-Trait Anxiety Inventory, TAT, and the Rorschach. Those that are specifically tailored toward aviation (e.g., the Cockpit Management Attitudes Questionnaire; Chidester, Helmreich, Gregorich, & Geis, 1991) do not focus on how pilots may make decisions in highly automated environments. Therefore, given the increasing automation technologies that are and will continue to be part of the cockpit and airspace system generally, scales should be developed to assess and measure the personality characteristics most suited for human-automation interaction in supervisory control situations.

**Training.** Related to this, most selection has taken place in the military aviation domain and the growing reduction in military pilots has increasingly led to *ab initio* training. Ab initio training is training for pilots from the very beginning rather than beginning commercial aviation training after logging x (hundreds) number of general aviation hours. Therefore, although Hunter and Burke (1995) stated that personality instruments may predict little for ab initio training performance, the impact of training ab initio pilots to safeguard against overreliance on automation is another promising approach to combat automation-induced complacency. Furthermore, repetition and experience are some of the greatest dangers to vigilance in the cockpit, such as the case of the very common complacent error of inadvertent wheels-up landings. It is so prevalent that pilots often say about it, "There are those that have and there are those that will." Recurrent training, such line-oriented-flight-training (LOFT), can be beneficial to help curb some of these "schema-driven" behaviors that can encourage the introduction of HSAs.

## ADAPTIVE AUTOMATION

These disadvantages of automation have resulted in increased interest in advanced automation concepts. One of these concepts is automation that is dynamic or adaptive in nature (Hancock & Chignell, 1987; Morrison, Gluckman, & Deaton, 1991; Rouse, 1977; 1988). In adaptive automation, control of tasks can be passed back and forth between the operator and automated systems in response to the changing task demands. Consequently, this allows for the restructuring of the task environment based

upon (a) what is automated, (b) when it should be automated, and (c) how it should be automated (Rouse, 1988; Scerbo, 1996). Rouse (1988) described the criteria for adaptive aiding systems:

The level of aiding, as well as the ways in which human and aid interact, should change as task demands vary. More specifically, the level of aiding should increase as task demands become such that human performance will unacceptably degrade without aiding. Further, the ways in which human and aid interact should become increasingly streamlined as task demands increase. Finally, it is quite likely that variations in level of aiding and modes of interaction will have to be initiated by the aid rather than by the human whose excess task demands have created a situation requiring aiding. The term *adaptive aiding* is used to denote aiding concepts that meet [these] requirements (p.432).

Adaptive aiding attempts to optimize the allocation of tasks by creating a mechanism for determining when tasks need to be automated (Morrison & Gluckman, 1994). In adaptive automation, the level or mode of automation can be modified in real-time. Further, unlike traditional forms of automation, both the system and the operator share control over changes in the state of automation (Scerbo, 1994; 1996). Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) have argued that adaptive automation represents the optimal coupling of the level of operator workload to the level of automation in the tasks. Thus, adaptive automation invokes automation only when task demands exceed the operator capabilities to perform the task(s) successfully. Otherwise, the operator retains manual control of the system functions. Although concerns have been raised about the dangers of adaptive automation (Billings & Woods, 1994; Wiener, 1989), it promises to regulate workload, bolster situational awareness, enhance vigilance, maintain manual skill levels, increase task involvement, and generally improve operator performance (Endsley, 1996; Parasuraman et al., 1992; Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1994, 1996; Singh, Molloy, & Parasuraman, 1993).

## **Research on Adaptive Technology**

During the 1970s, research on adaptive automation grew out of work in artificial intelligence. The direction of this effort was focused on developing the adaptive aids necessary to help determine human-automation task allocation between humans and on-board computers (Rouse, 1976; 1977). One program to emerge from this line of research was the Pilot's Associate; it was a joint effort among the Defense Advanced Research Projects Agency (DARPA), Lockheed Aeronautical Systems Company, McDonnell Aircraft Company, and the Wright Research and Development Center. The program sought to develop an "assistant" that could help provide the information necessary in an appropriate format to "assist" the pilot when they needed assistance. This was supplied through a network of cooperative knowledge-based subsystems that could monitor and assess events and then formulate plans to respond to problems (Hammer & Small, 1995). The work begun under the Pilot's Associate program was continued by the U.S. Army through the Rotorcraft Pilot's Associate (RPA) program (Colucci, 1995). The RPA program had similar goals as the Pilot Associate program in developing an intelligent "crew member" for the next generation of attack helicopters.

Recent research has continued the impressive work begun by the PA and RPA programs. Inagaki and his colleagues (1999, 2000) have explored the use of adaptive automation in managing go/no-go decisions during take-offs in commercial aircraft. The importance of this work is clear as the NTSB (1990) reported that pilots do not always make the correct decision under these circumstances. Inagaki, Takae and Moray (1999) have shown mathematically that the optimal approach to this problem is not one where the human pilot or the automation maintains full control over this decision. Rather, the best decisions are made when there is "shared control" over the decision-making process and this decision is made based upon critical factors such as actual airspeed, desired airspeed, the reliability of warnings, pilot

response time, etc. Inagaki et al. (1999) reported fewer errors were made when control over the decisions was traded between humans and the automation

Another example of applied research in the area of adaptive automation concerns the ever-present problem of Controlled Flight Into the Terrain (CFIT) that is one of the leading categories of accidents in commercial and military aviation (Khatwa & Roelen, 1996). A system is currently being tested, called the Ground Collision-Avoidance System (GCAS), for the F-16D that makes optimal determinations for terrain avoidance (Scott, 1999). The system provides a 5-sec warning to the pilot that terrain is present and threatens the aircraft (similar to GPWS warnings). If the pilot doesn't engage avoidance maneuvers within a certain timeframe, the GCAS presents an audio "fly up" warning and the GCAS takes control of the aircraft. Once the system has maneuvered the aircraft safely around the terrain, the system provides a message "You got it" and returns control of the aircraft to the pilot. Test pilots acknowledged the rapid intervention of the system and, when given the authority to override GCAS, eventually conceded control to the adaptive system.

**Adaptive Strategies.** Like much of technology, the technical potential is available but must be considered in terms of whether the technology should and ought to be implemented. Morrison and Gluckman (1994) reported on some early research that attempted to understand the dynamics of adaptive automation and how it should best be implemented. Strategies for invoking automation were based upon two primary factors: how they may be changed and what should be the "trigger" for making the change. Rouse and Rouse (1983), concerning the first point, described three different ways in which automation could assist the operator: 1. Whole tasks could be allocated to either the system or the operator to perform, 2. specific task(s) could be partitioned or divided so that the system and operator each share responsibility, or 3. task(s) could be changed into a different format or modality so as to manipulate the cognitive demands required to complete the task. Regarding the second point, a number of methods for adaptive automation have been proposed. Parasuraman and his colleagues (1992) reviewed the major techniques and found that they fell into five main categories: critical environmental events, operator performance measurement, operator modeling, physiological assessment, and hybrid methods. Recent research has been focused on the development of physiological assessment and hybrid methods (e.g., performance, operator models, and psychophysiology).

The physiological factors subelement has developed a research program on adaptive automation based on the use of psychophysiology for a number of reasons. A number of investigators in the area have noted that the best approach involves the assessment of measures that index the operators' state of mental engagement (Parasuraman et al., 1992; Rouse, 1988). The question, however, is what should be the "trigger" for the allocation of functions between the operator and the automation system. Numerous researchers have suggested that adaptive systems respond to variations in operator workload (Hancock & Chignell, 1987; 1988; Hancock, Chignell & Lowenthal, 1985; Humphrey & Kramer, 1994; Reising, 1985; Riley, 1985; Rouse, 1977), and that measures of workload be used to initiate changes in automation modes. Such measures include primary and secondary-task measures, subjective workload measures, and physiological measures. This, of course, presupposes that levels of operator workload can be specified so as to make changes in automation modes (Scerbo, 1996). Rouse (1977), for example, proposed a system for dynamic allocation of tasks based upon the operator's momentary workload level. Reising (1985) described a future cockpit in which pilot workload states are continuously monitored and functions are automatically reallocated back to the aircraft if workload levels get too high or too low. However, neither of these researchers provided specific parameters in which to make allocation changes (Parasuraman, 1990).

Morrison and Gluckman (1994), however, did suggest a number of workload indices candidates that may be used for initiating changes among levels of automation. They suggested that adaptive automation can be invoked through a combination of one or more real-time technological approaches. One of these proposed adaptive mechanisms is biopsychometrics. Under this method, physiological signals that reflect central nervous system activity, and perhaps changes in workload, would serve as a

trigger for shifting among modes or levels of automation (Hancock, Chignell, & Lowenthal, 1985; Morrison & Gluckman, 1994; Scerbo, 1996).

Byrne and Parasuraman (1996) discussed the theoretical framework for developing adaptive automation around psychophysiological measures. The use of physiological measures in adaptive systems is based on the idea that there exists an optimal state of engagement (Gaillard, 1993; Hockey, Coles, & Gaillard, 1986). Capacity and resource theories (Kahneman, 1973; Wickens, 1984; 1992) are central to this idea. These theories posit that there exists a limited amount of resources to draw upon when performing tasks. These resources are not directly observable, but instead are hypothetical constructs. Kahneman (1973) conceptualized resources as being limited, and that the limitation is a function of the level of arousal. Changes in arousal and the concomitant changes in resource capacity are thought to be controlled by feedback from other ongoing activities. An increase in the activities (i.e., task load) causes a rise in arousal and a subsequent decrease in capacity. Kahneman's model was derived from research (Kahneman et al., 1967, 1968, 1969) on pupil diameter and task difficulty. Therefore, physiological measures have been posited to index the utilization of cognitive resources.

Several biopsychometrics have been shown to be sensitive to changes in operator workload suggesting them as potential candidates for adaptive automation. These include heart rate variability (Bucks, Ryan, & Wilson, 1994; Itoh, Hayashi, Tsukui, & Saito, 1989; Lindholm & Cheatham, 1983; Lindqvist et al., 1983; Opmeer & Krol, 1973; Sayers, 1973; Sekiguchi et al., 1978), EEG (Natani & Gomer, 1981; O'Hanlon & Beatty, 1977; Sterman, Schummer, Dushenko, & Smith, 1987; Torsvall & Akerstedt, 1987), eyeblinks (Goldstein, Walrath, Stern, & Strock, 1985; Sirevaag, Kramer, deJong, & Mecklinger, 1988), pupil diameter (Beatty, 1982; 1986; 1988; Qiyuan, Richer, Wagoner, & Beatty, 1985; Richer & Beatty, 1985; 1987; Richer, Silverman, & Beatty, 1983), electrodermal activity (Straube et al., 1987; Vossel & Rossmann, 1984; Wilson, 1987; Wilson & Graham, 1989) and event-related potentials (Defayolle, Dinand, & Gentil, 1971; Gomer, 1981; Hancock, Chignell, & Lowenthal, 1985; Reising, 1985; Rouse, 1977; Sem-Jacobson, 1981).

The advantage to biopsychometrics in adaptive systems is that the measures can be obtained continuously with little intrusion (Eggemeier, 1988; Kramer, 1991; Wilson & Eggemeier, 1991). Also, because behavior is often at a low level when humans interact with automated systems, it is difficult to measure resource capacity with performance indices. Furthermore, these measures have been found to be diagnostic of multiple levels of arousal, attention, and workload. Therefore, it seems reasonable to determine the efficacy of using psychophysiological measures to allocate functions in an adaptive automated system. However, although many proposals concerning the use of psychophysiological measures in adaptive systems have been advanced, not much research has actually been reported (Byrne & Parasuraman, 1996). Nonetheless, many researchers have suggested that perhaps the two most promising psychophysiological indices for adaptive automation are the electroencephalogram (EEG) and event-related potential (ERP) (Byrne & Parasurman, 1996; Kramer, Trejo, & Humphrey, 1996; Morrison & Gluckman, 1994; Parasuraman, 1990; Scerbo, 1996). In addition to heart-rate metrics (e.g., heart-rate interbeat interval and variability) that are the focus of fiscal year 2001-2002 research; there has been other significant amounts of research examining the use of EEG and ERPs for adaptive automation design.

## **Electroencephalogram**

To provide some background, the following short section details the physiological basis of the EEG and gives an abbreviated description of research that has shown its promise in mental workload assessment. For more information on EEG and adaptive automation, the interested reader is pointed to a NASA publication that was a collaborative effort between NASA Langley Research Center, Old Dominion University, and Catholic University, that examined the use of these measures for adaptive automation design (Scerbo et al., 2001).

**Physiological Basis.** The EEG derives from activity in neural tissue located in the cerebral cortex, but the precise origin of the EEG, what it represents, and the functions that it serves are not presently known. Current theory suggests that the EEG originates from post synaptic potentials rather than action potentials. Thus, the EEG is postulated to result primarily from the subthreshold post-synaptic potentials that may summate and reflect stimulus intensity instead of firing in an all-or-none fashion (Gale & Edwards, 1983).

**Description of the EEG.** The EEG consists of a spectrum of frequencies between 0.5 Hz to 35 Hz (Surwillo, 1990). Delta waves are large amplitude, low frequency waveforms that typically range between 0.5 and 3.5 Hz in frequency, in the range of 20 to 200  $\mu$ V (Andreassi, 1995). Theta waves are a relatively uncommon type of brain rhythm that occurs between 4 and 7 Hz at an amplitude ranging from 20 to 100  $\mu$ V. Alpha waves occur between 8 and 13 Hz at a magnitude of 20 to 60  $\mu$ V. Finally, beta waves are an irregular waveform at a frequency of 14 to 30 Hz at an amplitude of about 2 to 20  $\mu$ V (Andreassi, 1995). An alert person performing a very demanding task tends to exhibit predominately low amplitude, high Hz waveforms (beta activity). An awake, but less alert person shows a higher amplitude, slower frequency of activity (alpha activity). With drowsiness, theta waves predominate and in the early cycles of deep slow wave sleep, delta waves are evident in the EEG waveform. The generalized effect of stress, activation or attention is a shift towards the faster frequencies, lower amplitudes with an abrupt blocking of alpha activity (Horst, 1987).

**Laboratory Studies.** Gale (1987) found that there exists an inverse relationship between alpha power and task difficulty. Other studies have also demonstrated the sensitivity of alpha waves to variations in workload associated with task performance. Natani and Gomer (1981) found decreased alpha and theta power when high workload conditions were introduced to pilots during pitch and roll disturbances in flight. Serman, Schummer, Dushenko, and Smith (1987) conducted a series of aircraft and flight simulation experiments in which they also demonstrated decreased alpha power and tracking performance in flight with increasing task difficulty.

Numerous studies have also demonstrated that theta may be sensitive to increases in mental workload. Subjects have been trained to produce EEG theta patterns to regulate degrees of attention (Beatty, Greenberg, Diebler, & O'Hanlon, 1974; Beatty & O'Hanlon, 1979; O'Hanlon & Beatty, 1979; O'Hanlon, Royal, & Beatty, 1977). In particular, Beatty and O'Hanlon (1979) found that both college students and trained radar operators, who had been taught to suppress theta activity performed better than controls on a vigilance task. Though theta regulation has been shown to affect attention, the magnitude of the effect is often small (Alluisi, Coates, & Morgan, 1977). More recent research, however, has demonstrated its utility in assessing mental workload. Both Natani and Gomer (1981) and Sirevaag, Kramer, deJong, and Mecklinger (1988) found decreases in theta activity as task difficulty increased and during transitions from single to multiple tasks, respectively.

**Field Research.** More recent research has demonstrated the utility of EEG in assessing mental workload in the operational environment. Serman et al. (1993) evaluated EEG data obtained from 15 Air Force pilots during air refueling and landing exercises performed in an advanced technology aircraft simulator. They found a progressive suppression of 8-12 Hz activity (alpha waves) at medial (Pz) and right parietal (P4) sites with increasing amounts of workload. Additionally, a significant decrease in the total EEG power (progressive engagement) was found at P4 during the aircraft turning condition for the air refueling task (the most difficult flight maneuver). This confirmed other research that found alpha rhythm suppression as a function of increased mental workload (e.g., Ray & Cole, 1985).

## The Biocybernetic System

The Crew Hazards & Error Management (CHEM) group, that is now the team of the physiological factors element of Physiological / Psychological Stressors and Factors program at NASA Langley Research Center, had developed a biocybernetic closed-loop system for the investigation of

physiological measures for adaptive automation. Pope, Bogart, and Bartolome (1995) first reported on the system and is one of the few studies examining the utility of EEG for adaptive automation technology. These researchers developed an adaptive system that uses a closed-loop method to adjust modes of automation based upon changes in the operator's EEG patterns. The closed-loop method was developed to determine optimal task allocation using an EEG-based index of engagement or arousal. The system uses a biocybernetic loop that is formed by changing levels of automation in response to changing taskload demands. These changes were made based upon an inverse relationship between the level of automation in the task set and the level of pilot workload.

The level of automation in a task set could be such that all, none, or a subset of the tasks could be automated. The task mix is modified in real time according to operator's level of engagement. The system assigns additional tasks to the operator when the EEG reflects a reduction in task set engagement. On the other hand, when the EEG indicates an increase in mental workload, a task or set of tasks may be automated, reducing the demands on the operator. Thus, the feedback system should eventually reach a steady-state condition in which neither sustained rises nor sustained declines in the EEG are observed. Figure 6 presents a graphic of the system.

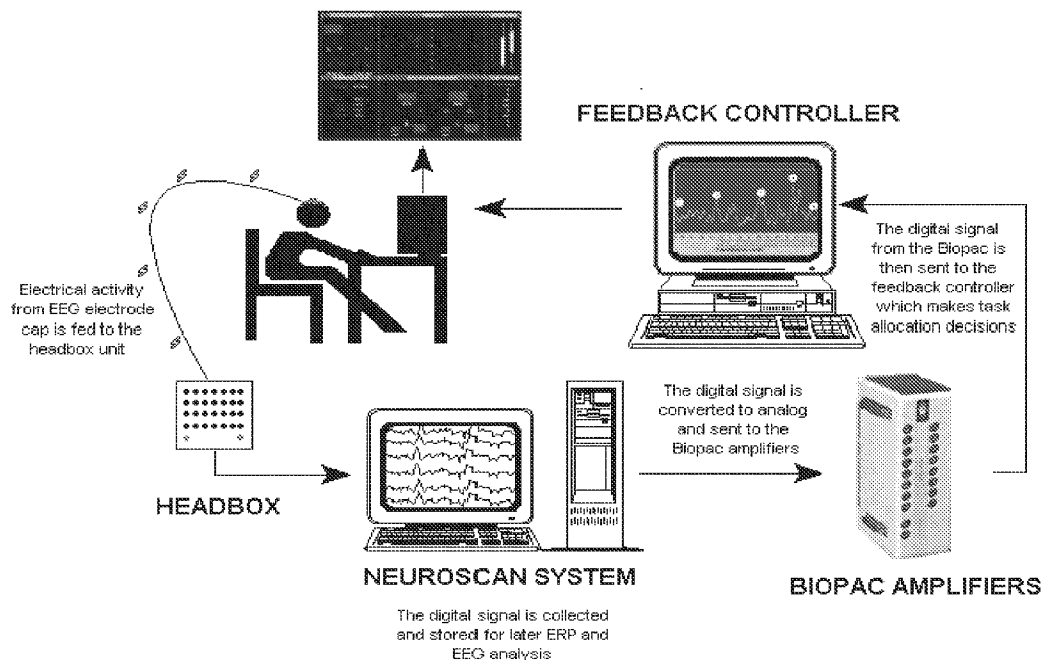


Figure 6. The Biocybernetic Closed-Loop System

One issue for the biocybernetic system concerns the nature of the EEG signal used to drive changes in task mode. We argued that differences in task demand elicit different degrees of mental engagement that could be measured through the use of EEG-based engagement indices. These researchers tested several candidate indices of engagement derived from EEG power bands (alpha, beta, & theta). These indices of engagement were derived from recent research in vigilance and attention (Davidson, 1988; Davidson et al., 1990; Lubar, 1991; Offenloch & Zahner, 1990; Streitberg, Rohmel, Herrmann, & Kubicki, 1987). For example, Davidson et al. (1990) argued that alpha power and beta

power are negatively correlated with each other to different levels of arousal. Therefore, these power bands can be coupled to provide an index of arousal. For example, Lubar (1991) found that the band ratio of beta/theta was able to discriminate between normal children and those with attention deficit disorder.

We have reasoned that the usefulness of a task engagement index would be determined by a demonstrated functional relationship between the candidate index and task operating modes (i.e., manual versus automatic) in the closed-loop configuration. We have used both positive and negative feedback controls to test candidate indices of engagement because each should impact system functioning in the opposite way, and a good index should be able to discriminate between them. For example, under negative feedback conditions, the level of automation in the tasks was lowered (i.e., automated) when the EEG index reflected increasing engagement. On the other hand, when the EEG reflected increases in task demands, automation levels were increased. Task changes were made in the opposite direction under positive feedback conditions; that is, the level of automation in the tasks was maintained when the EEG engagement index reflected increasing task demands. If there was a functional relationship between an index and task mode, the index should demonstrate stable short-cycle oscillation under negative feedback and longer and more variable periods of oscillation under positive feedback. The strength of the relationship would be reflected in the degree of contrast between the behavior of the index under the two feedback contingencies.

Our original findings were that that the closed-loop system was capable of regulating participants' engagement levels based upon their EEG activity. The index beta/(alpha+theta) possessed the best responsiveness for discriminating between the positive and negative feedback conditions (see Figure 7). The conclusion was based upon the increased task allocations in the negative feedback condition witnessed under this index than under either the beta/alpha or alpha/alpha indexes. These results were taken to suggest that the closed-loop system provides a means for evaluating the use of psychophysiological measures for adapting automation.

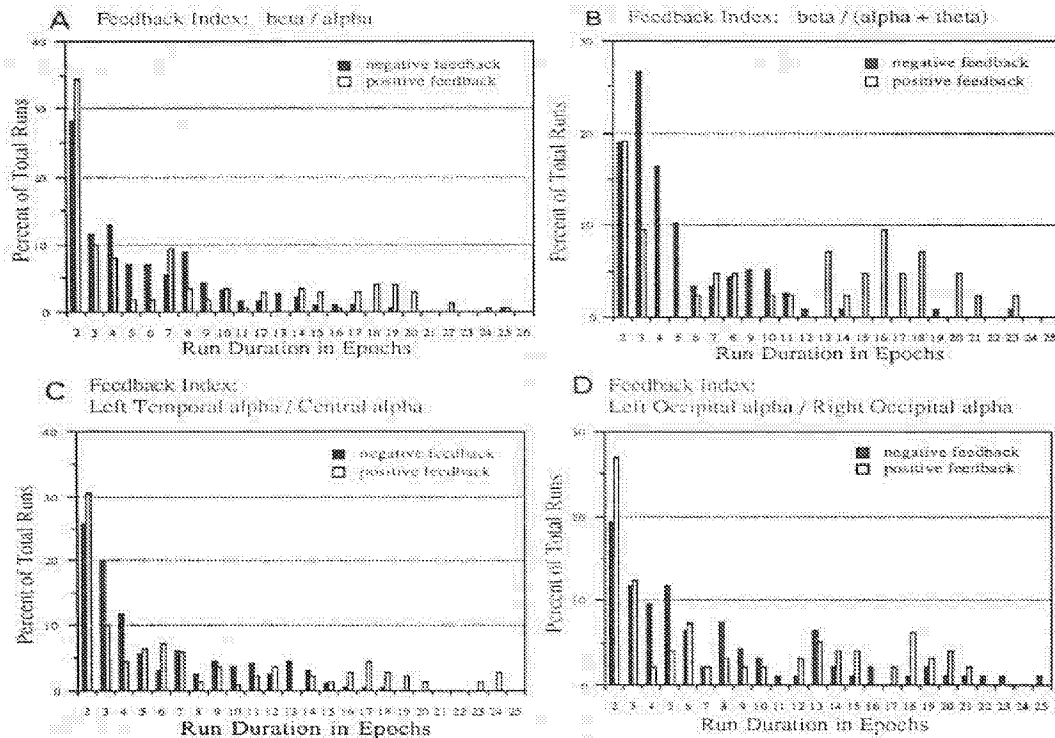


Figure 7. The Run Distributions for each Candidate Index.



The system operates with a moving window procedure in which the EEG is recorded for 40 seconds to determine the initial value of the engagement index. The window is advanced 2 seconds and a new average is derived. Originally, we argued that an increase in the slope of two consecutive values of the index reflects an increase in engagement (i.e., increasing slope), and a decreasing slope reflected a decrease in engagement.

Later studies have expanded on the original studies and have focused on the performance outcomes of using the biocybernetic system to control automation task allocation. After all, adaptive automation has the intended benefits of improving pilot performance, lowering workload, and improving situation awareness. The original studies did not focus on the performance or physiological data since the intention was to demonstrate feedforward state contingent behavior of the EEG engagement indices. Follow-up studies, however, examined the use of the indices across a large number of subjects and examined the association of the index behavior with performance, workload, and psychophysiological data. As an example, a study was conducted, under the physiological factors element, which examined system operation with the indices of  $1/\alpha$ ,  $\beta/\alpha$ , and  $\beta/(\alpha+\theta)$  gathered from Cz, Pz, P3, and P4. Furthermore, performance was assessed based on root-mean-squared-error (RMSE) of the tracking scores.



Figure 8. Research on Adaptive Automation at NASA Langley Research Center

As in our first studies, there was a main effect for feedback condition,  $F(1,15) = 7.34$ ,  $p < .02$  with more task allocations under the negative feedback condition than under the positive feedback condition. Furthermore, there were more task allocations made under the  $\beta/(\alpha+\theta)$  index than the other two EEG engagement indices and the interaction,  $F(2,15) = 5.25$ ,  $p < .02$  showed that these task allocations were confined to the negative feedback condition as hypothesized. This finding supported our previous results in helping to valid system operation according to cybernetic feedback contingencies.

Additionally, participants had lower RMSE scores under the negative feedback condition,  $F(1,15) = 25.02$ ,  $p < .001$  and the EEG discriminated feedback contingency and automation mode (see Table 5).

**Table 5.** Mean Values of EEG Engagement Index (beta/alpha+theta) for Task Mode and Feedback Conditions

	Task Mode	
	Automated	Manual
Negative Feedback	31	21
Positive Feedback	20	30

Collectively, with other similar studies, the results help to validate the operation of the biocybernetic system and to demonstrate that negative feedback control of automation mode based on EEG engagement index can significantly improve performance and increase task engagement.

## WORKLOAD

A second series of studies were conducted to examine the effect that system operation has on operator workload. The experiments used a similar system to examine the effectiveness of the EEG engagement index, beta / (alpha+theta), to produce expected feedback control behavior. Thus, the value of the index was expected to oscillate in a more regular and stable pattern under negative feedback than under positive feedback. Consequently, more task allocations were expected under negative feedback than under positive feedback. Furthermore, our past results were generated using only a single compensatory tracking task. Therefore, the objective of these studies were to examine system operation under both single- and multiple-task conditions. Multiple Resource Theory (Wickens, 1984, 1992) posits that performance on a task that is performed in conjunction with other tasks should be poorer than performance on a task performed alone because of competition for cognitive resources (if they compete for the same resources). For example, Parasuraman, Molloy, and Singh (1993) asked participants to perform either a system monitoring task (single task condition) or a system monitoring, compensatory tracking, and a resource management task (multiple task condition). Participants missed fewer critical signals while performing the system-monitoring task alone than when performing all the tasks concurrently. Results such as these are not limited to monitoring tasks. For example, Arnegard (1991) found that the combination of these same tasks resulted in significant increases in workload compared to only the compensatory tracking task. The results of these studies suggest that multiple task conditions produce higher levels of workload and can lead to decreases in performance.

Automation-induced performance decrements in multiple task performance may stem from changes in the processing strategies that participants use to devote cognitive resources to the different tasks. A number of researchers have stated that operators may become complacent as they gain more experience with automation leading to an increase in trust and reliance on automation (Riley, 1994; Singh et al., 1993). Such shifts in strategy do not provide adequate processing resources for the maintenance of automated tasks. It has been suggested that adaptive systems, however, are less susceptible to automation-induced performance decrements because of the regulation of workload and maintenance of operator engagement (Hancock & Chignell, 1988; Scerbo, 1996). The closed-loop system was designed to moderate workload by reducing task demands when levels of workload increase. Accordingly, we expected that the biocybernetic system would make more task allocations under the multiple task condition to compensate for the increased fluctuations in task load that would accompany the operation of multiple tasks, each with their own unique demand schedules. Furthermore, performance under the

multiple task condition was predicted to be significantly better for participants who performed these tasks under the closed-loop system than a control group who performed these tasks without the benefit of adaptive task allocation.

As an example of one of these research studies, we asked 40 participants to perform either the tracking task (single task condition) or the tracking task, system monitoring task, and resource management task (multiple task condition). The design was a 2 (feedback condition; positive or negative feedback) X 2 (task mode, automatic or manual) X 2 (task level; single or multiple task condition) X 2 (group; experimental or control group) mixed-subjects design. The system was similar to our previous studies and, therefore, made task allocation decision based on the behavior of the EEG engagement index. Our results supported the conclusion that the system was able to moderate workload demands by regulating the level of human-automation mix in order to optimize operator performance. Participants did significantly better,  $F(1,23) = 78.57$  under the negative feedback condition overall and that participants, that were subjected to adaptive task allocation, did significantly better than the control group (who received static automation),  $F(1,47) = 4.049$ ,  $p < .05$ . Moreover, these participants rated subjective workload, from the NASA-TLX, to be significantly lower,  $F(1,23) = 46.05$ ,  $p < .001$ . Such results point that the system does indeed moderate workload and task engagement by demonstrating significant improvements in subject performance and reported workload even under conditions of high workload.

**Table 6. Mean Values for Dependent Physiological Measures**

EEG Measures	Manual				Automatic			
	Negative		Positive		Negative		Positive	
	M	SD	M	SD	M	SD	M	SD
Engagement index	8.689	3.423	11.739	4.944	11.513	4.523	8.675	3.577
Alpha	0.047	0.029	0.041	0.022	0.043	0.025	0.052	0.033
Beta	0.034	0.016	0.040	0.019	0.039	0.019	0.036	0.017
Theta	0.047	0.022	0.042	0.019	0.041	0.018	0.049	0.021

*Note.* Values represent mean relative power.

## Research on Enhancements to Biocybernetic System

Research has been conducted, under the physiological factors element, to examine various parameters of the biocybernetic system to better able the regulation of pilot engagement and workload through the use of adaptive task allocation. The earlier version of the system used a moving window that calculates a *slope* that reflects increasing or decreasing engagement. Studies were run to see the impact of changing how the system made task allocation decisions and these focused on changing aspects of the moving window derivation of the EEG engagement index. The first of these studies is provided through the example below that was conducted to calculate the optimal window size for deriving the EEG engagement index. Afterwards, a second example study is presented that focused on a new derivation procedure that relies on an absolute method of calculating the EEG engagement index based on baseline mean and standard deviation statistics.

**Moving Window Size.** The present study varied the length of the EEG window used to drive the system. The system computes an EEG index every 2 seconds (one epoch) and the derived arousal index was updated using either a sliding 4 second window (2 epochs) or using a sliding 20 second window (10 epochs). The smaller window was expected to produce a more responsive system (more task allocations or switches). A system that is more responsive to operator arousal is expected to produce better performance.

The subjects were 14 undergraduate students (males & females) between the ages of 18 and 50. The Multiple Attribute Task Battery was used. In this study the subjects performed the compensatory tracking task and the monitoring task of the MAT, but only the tracking task was controlled by the adaptive automation system.

Subjects were run under both the 2 and 10 epoch windows (4 and 20 seconds) and under positive and negative feedback conditions. Upon arriving at the laboratory each subject was explained the nature of the task and fitted with the electrode cap. The left mastoid area was used as the reference with four recording sites (Cz, Pz, P3, and P4). Each subject then performed the tracking task for the five minute baseline. After this the first of two 16-minute tracking sessions began. Half of the subjects were run under the 2 epoch window and half under the 10 epoch window. Each 16- minute session consisted of alternating four minute positive and negative feedback conditions. After completing the first session the subjects were given a five minute rest and then received another 5 minute baseline. The epoch condition for each subject was reversed for the second session.

The initial analysis for this study assessed whether the system was operating correctly and switching subjects on the basis of the derived EEG-index. To this end, an Epochs (2, 10) X Feedback (negative, positive) X Task (automatic, manual) analysis of variance of the arousal index was performed. Marginal main effects for Feedback and Task were found,  $F_s(1, 11) = 4.26$  and  $3.87$ ,  $p < .10$ , respectively. The interaction of Feedback and Task was significant,  $F(1, 11) = 7.90$ ,  $p < .02$ . The pattern of the EEG index indicates that there were higher levels of arousal under positive feedback coupled with the manual condition and the negative feedback coupled with automatic condition. The pattern is repeated in the significant Feedback X Task X Epoch interaction,  $F(1, 11) = 5.66$ ,  $p < .04$ . A Tukey Post Hoc analysis of this interaction indicated that for the short 2 epoch condition there were significant differences between the manual and automatic indexes for both feedback conditions. A higher index under positive feedback was found for the manual tracking condition, whereas under the negative feedback condition a higher index occurred in the automatic condition. For the 10 epoch condition the same pattern occurred with the exception that the difference under negative feedback was not significant, although there was a trend for a higher index under the automatic condition. There were no other significant effects in this analysis.

A mixed design Epochs X Feedback analysis of variance was performed on the number of switches. This analysis yielded a significant main effect for Epoch,  $F(1, 10) = 40.11$ ,  $p < .0001$  with more switches occurring under the shorter 2-epoch condition than the 10-epoch condition (42.4 and 15.9, respectively). Also, the main effect of Feedback was significant,  $F(1, 13) = 7.28$ ,  $p < .02$ , with more switching under the negative feedback than the positive feedback condition (33.4 and 23.9, respectively).

A similar analysis was conducted on the tracking error as a function of feedback conditions and epoch levels. This analysis yielded a main effect for Epoch,  $F(1, 12) = 10.32$ ,  $p < .01$  with better tracking (less error) under the 2-epoch than the 10-epoch condition (13.1 and 17.9, respectively). Also, the main effect of Feedback was significant,  $F(1, 12) = 14.47$ ,  $p < .005$ , with better tracking under the negative feedback condition (13.5 and 17.5, respectively). The interaction of these conditions was marginally significant,  $F(1, 12) = 3.21$ ,  $p < .10$ . An exploratory Tukey HSD test on the four means indicated that tracking error under the negative feedback and the 2 epoch condition was less than all other conditions which were not different from each other.

The primary finding of this research on an adaptive automation system was that the system performed as designed and shifted task functions appropriately depending on the operator's state of arousal and the feedback contingency. Essentially, higher indexes were found under the positive manual and the negative automatic conditions. The second finding was that the system produced many more task switches under the negative feedback condition, which is consistent with the notion of keeping the subjects at a stable level of arousal (Pope, et. al. 1995). Further, the use of a shorter epoch window for deriving the EEG index resulted in more switching. The critical finding of the study was that tracking performance was better under both the short epoch window and the negative feedback condition. The value of an adaptive automation system must be in maintaining or improving performance. These findings indicate that procedures employed in the present study may be used to evaluate the parameters important to a functional adaptive

automation system. Also, the on-line assessment of mental state as defined by the EEG index may be used to make task allocation decisions. One avenue of further investigation should be on the relative effectiveness of various cortical sites for the derived EEG-index. Additionally, we need to examine this system using the full battery of tasks, varying which task are controlled by adaptive automation, and using much longer performance periods. Lastly, it is possible to use the system to define index "steps" and make finer gradations in task allocations rather than just a manual-automatic dichotomy. With complex systems the number of tasks automated and the level of automation of each may be varied as a function of the EEG-derived index.

**Absolute Index.** Three experiments were conducted to evaluate the performance of the system using an absolute criterion to make task allocation decisions. The slope method used the successively derived (every 2 sec.) engagement index to establish the slope and, more importantly, the sign, positive or negative, which determined the task state, automated or manual. Under the negative feedback condition, the task was switched to or remained in manual mode when the index slope was decreasing (negative slope) and switched to or remained in automatic mode when the index was increasing (positive slope). Under the positive feedback condition, the task was switched to or remained in manual mode when the index slope was increasing and switched to or remained in automatic mode when the index was decreasing. To contrast, the absolute method used a 5-min baseline prior to testing, and the mean engagement index of this period served as the engagement index threshold for task state, with values above and below the threshold controlling the switches. Under the negative feedback condition, the task was switched to or remained in manual mode when the index was below the baseline index and switched to or remained in automatic mode when the index was above the baseline index. Under the positive feedback condition, the task was switched to or remained in manual mode when the index was above the baseline index and switched to or remained in automatic mode when the index was below the baseline. Comparing the absolute method to the slope method, there was a significant effect found for performance with negative feedback producing significantly less tracking error than positive feedback,  $t(11) = 2.27$ ,  $p < 2.27$ . However, overall performance was better using the slope method of calculation than the absolute method although the interaction was not significant,  $p < .05$  (see Figure 9). Given the problem of individual differences in psychophysiology, therefore, the use of the absolute method was determined to be the better choice and was adopted in future studies employing the closed-loop, biocybernetic system.

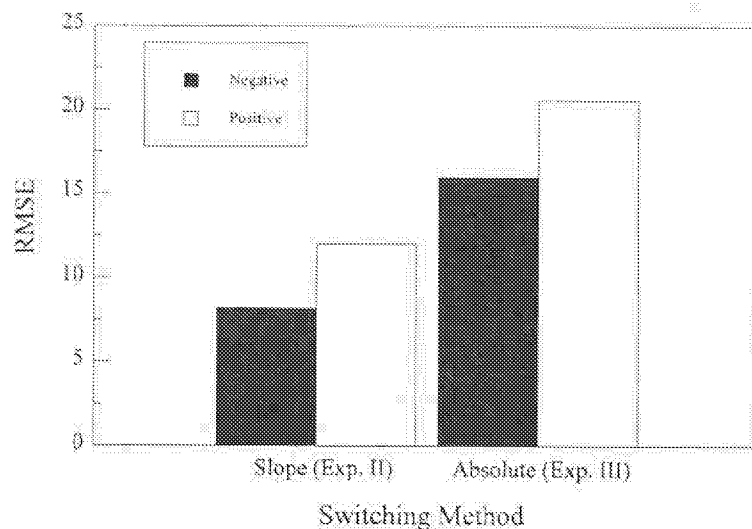


Figure 9. Absolute versus Slope Method for EEG Engagement Index

## Supervisory Control Monitoring

The closed-loop system and research conducted on adaptive automation had been focused on changing the difficulty or task modes of a compensatory tracking task. The task can be classified as a motor task and we were interested in how the system could benefit other types of sensory and/or cognitive processes. Therefore, a study was conducted that adaptively changed aspects of a vigilance-type perceptual task based on whether the subject was engaged in the task or not.

The goal of the study was to determine whether the improvements in tracking performance observed with a biocybernetic, closed loop system would extend to vigilance performance. Specifically, an index of engagement was used to change event rates and therefore moderate performance. Two groups were under different patterns of feedback. Under negative feedback conditions, an increase in engagement produced a decrease in event rate and lower levels of engagement resulted in an increase in event rate. The opposite pattern of changes occurred in positive feedback conditions. It was expected that under negative feedback conditions, that lowering the event rate when engagement levels were high and increasing event rate when engagement declined would serve to stabilize vigilance performance and therefore either eliminate or attenuate the decrement. By contrast, positive feedback conditions would drive event rate and engagement levels to the extremes and therefore result in the expected vigilance decrement.

**Vigilance Studies.** One could argue that there might be benefits to the unique pattern of changes among event rates that would facilitate performance irrespective of the value of the engagement index. 40 participants in negative and positive feedback conditions were yoked to other participants who received the same patterns of switches between automatic and manual task modes, but whose EEG signals did not affect the system. Previous research in our lab has found that the experimental subjects in the negative feedback condition had better tracking performance than their yoked counterparts, and both performed better than the positive feedback groups. Thus, two additional groups were run in the present study. These groups were yoked to the negative and positive feedback groups. They received the same exact patterns of event rate changes as their counterparts in the two conditions described previously, but the engagement levels derived from their own EEG signals had no effect on the display. These yoked control groups were included to determine whether an individual's own engagement index-generated pattern or a pattern generated by another participant were responsible for effects observed in the present study.

The task was a 40-min vigilance task which required participants to monitor the repetitive presentation of a pair of 3mm (W) X 38mm (H) white lines separated by 25mm. These lines appeared in the center of the monitor screen. The stimuli were white and were presented on a blue background. Critical signals (targets) were 2mm taller and occurred once a minute at random intervals. The participants were required to respond to the presence of the critical signals by pressing the space bar on the keyboard.

Three different event rates were used (6, 20, and 60 events per minute). All sessions began with a 5-min baseline period with the event rate set at 20. The mean and standard deviation of the engagement index was derived from the baseline. This value was then used to change the event rate when the index fell or rose .2 sd or more above the baseline index. For the experimental participants in the negative feedback condition, the event rate increased to 60 when the engagement index dropped .2 sd below the baseline value and decreased to 6 when the engagement index rose .2 sd above baseline. Conversely, for experimental participants in the positive feedback condition the event rate increased to 60 when the engagement index rose .2 sd above the baseline value and decreased to 6 when the engagement index fell .2 sd below the baseline value. Each participant in both feedback groups was paired with a yoked-control participant. These participants received the same pattern of stimulus changes as their experimental partner. The EEG signals of the yoked-control participants were also recorded; however, in the yoked

conditions the engagement index had no effect on the changes in the stimulus event rate presented to them.

A 2 Feedback (positive, negative) by 2 Control Condition (experimental, yoked) by 4 Periods ANOVA of the A' scores yielded significant main effects for Feedback,  $F(1, 36) = 9.23$ ,  $p < .004$ , and Periods,  $F(3, 108) = 7.04$ ,  $p < .002$ . These main effects reflect the higher levels of vigilance sensitivity for the negative feedback condition and the general decline in vigilance over the four periods. The positive feedback condition produced a clear decline in performance over periods, but there was little change in the negative feedback condition. Exploratory Tukey HSD comparisons for these two conditions indicated that the positive feedback vigilance scores were lower on the third and fourth periods compared to the first period. There were no comparable differences in the negative feedback condition. A comparison between the conditions indicated that they were equivalent in the first period, but the positive feedback scores were significantly lower over the final three periods.

Importantly, the main effect for control condition and the accompanying interaction terms indicated that there were no differences between the yoked groups and their respective experimental groups. A similar Feedback by Control Conditions by Periods analyses of the B' scores did not yield any significant findings. There appeared to be no difference in the decision criteria used by the four experimental groups. A 2 Feedback by 4 Periods ANOVA was performed on the mean event rates generated in the experimental conditions. The results yielded a significant effect for Feedback,  $F(1, 18) = 9.12$ ,  $p < .008$ , and a marginal Feedback by Periods interaction,  $F(3, 54) = 3.07$ ,  $p < .08$ . The nature of the Feedback by Periods interaction is shown in Figure 10. Tukey HSD comparisons of the interaction revealed that the feedback conditions were not different during the first period, but the event rates of the negative feedback condition were significantly lower over the final three periods.

Correlations were computed between the A' scores and event rates for each period for each condition separately (positive feedback, negative feedback). None of the correlations for the negative feedback condition approached significance. For the positive feedback condition the correlations for the three periods were negative but only reached significance for the last period,  $r(18) = -0.51$ , which indicates that vigilance performance was poorer with higher event rates. Eta-square estimations of the relative variance related to each F-ratio in this analysis indicated that the Feedback by Event Rate interaction accounted for 24 percent of the variance, while none of the other interactions accounted for more than 2 percent of the variance. Of the significant interaction terms, none accounted for more than 0.1 percent of the variance.

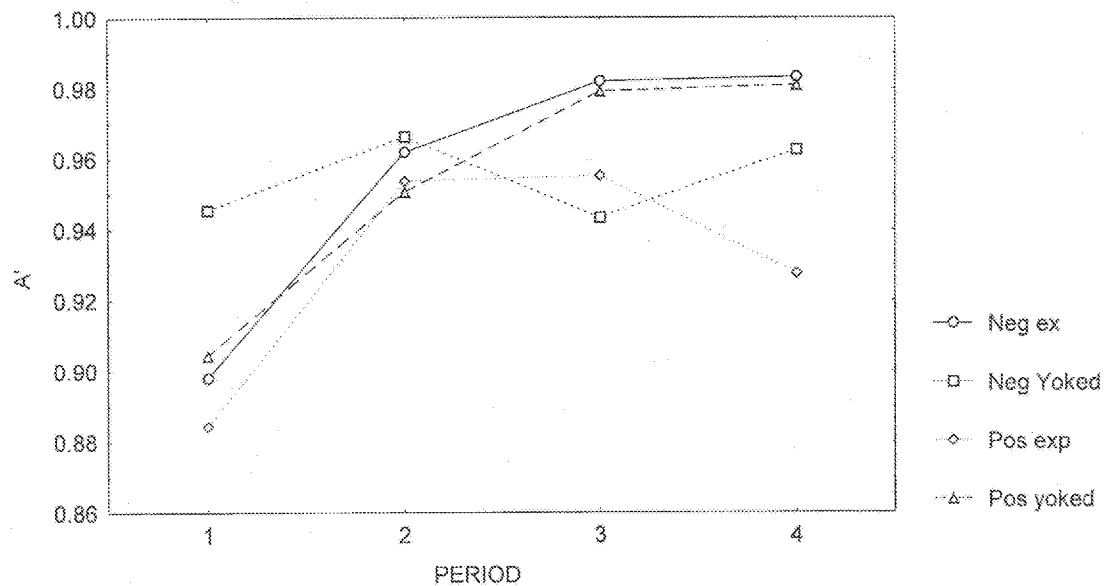


Figure 10. Period X Experimental Group Interaction for A'

The results from the present study are important for two reasons. First, the findings indicate that a biocybernetic, closed loop system using an EEG index of engagement may facilitate types of performance other than the psychomotor tracking activities for which it was originally designed. Second, to our knowledge the results from the present study represent one of the few experimental manipulations to eliminate the vigilance decrement. The idea that a pattern of event rate changes can facilitate monitoring performance could be established without any overt action on the part of the observer is intriguing indeed. Additional research will be needed to examine more closely those task parameters that affect vigilance performance moderated by one's own EEG.

## Event-Related Potentials and Adaptive Automation

The EEG and ERP represent viable candidates for determining shifts between modes of automation in adaptive systems. Because real-time assessment of workload is the goal of system designers wanting to implement adaptive automation, it is likely that these measures will become the focus of research on adaptive automation. This optimism stems from a number of studies that have suggested that they might be useful for on-line evaluations of operator workload (Defayolle et al., 1971; Farwell & Donchin, 1988; Gomer, 1981; Humphrey & Kramer, 1994; Kramer, 1991; Kramer et al., 1989; Sem-Jacobsen, 1981). Although these results suggest that on-line assessment of mental workload may be possible in the near future, a good deal of additional research is needed.

The determination of measures on which to dynamically allocate automation does not represent the only area that needs exploration. Other areas include the frequency with which task allocations are made, when automation should be invoked, and how this invocation changes the nature of the operator's task (Parasuraman et al., 1992). Specifically, it is not known how changing among automation task modes impacts the human-automation interaction. Therefore, one of the studies conducted in the physiological factors element attempted to examine the efficacy of use of EEG and ERPs for adaptive task allocation was also examined.

The use of ERPs in the design of adaptive automation systems was considered some years ago in the context of developing "biocybernetic" communication between the pilot and the aircraft (Donchin,



1980; Gomer, 1981). The idea concerned systems in which tasks or functions could be allocated flexibly to operators, using ERPs, which may allow the optimization of mental workload to be sought in a dynamic, real-time environment. For example, a method might be developed for obtaining momentary workload levels allowing an index to be derived, such as the amplitude of the P300 wave of the ERP. The workload index could then be compared in real-time to a stored profile of the ERP associated with that task(s). The profile would be generated from initial baseline data. If the optimal physiological level for a task is exceeded, then the task(s) could be off-loaded from the operator and allocated to the system. Further, if the workload levels become too low, then the task(s) could be transferred back to the operator (Parasuraman, 1990). In recent reviews, however, Parasuraman (Byrne & Parasuraman, 1996; Parasuraman, 1990) concluded that although many proposals have been made concerning the use of ERPs in adaptive systems, little actual research has been conducted.

The proposed study attempted to further the research on the use of ERPs for adaptive automation. What is proposed is that the absolute biocybernetic system be used to make task allocation decisions between manual and automatic task modes as previously described. Participants were also asked to perform an oddball, auditory task concurrently with the compensatory tracking task. The EEG signal was fed to both the biocybernetic system and to a data acquisition system that permitted the analysis of ERPs to high and low frequency tones. Such results are hoped to assess the efficacy of using ERPs in the design of adaptive automation technology.

Thirty-six participants were randomly assigned to an experimental, yoked, or control group condition. Under the experimental condition, a compensatory tracking task was switched between manual and automatic task modes based upon the participant's EEG. ERPs were also gathered to an auditory, oddball task. Participants in the yoked condition performed the same tasks under the exact sequence of task allocations that participants in the experimental group experienced. The control condition consisted of a random sequence of task allocations that was representative of each participant in the experimental group condition. Therefore, the design allowed a test of whether the performance and workload benefits seen in previous studies using the biocybernetic system were due to adaptive aiding or merely to the increase in task mode allocations.

The results showed that the use of adaptive aiding improved performance and lowered subjective workload under negative feedback as predicted. Additionally, participants in the adaptive group had significantly lower tracking error scores and NASA-TLX ratings than participants in either the yoked or control group conditions (Figures 11 & 12). Furthermore, the amplitudes of the N1 and P3 ERP components were significantly larger under the experimental group condition than under either the yoked or control group conditions (see Figure 13).

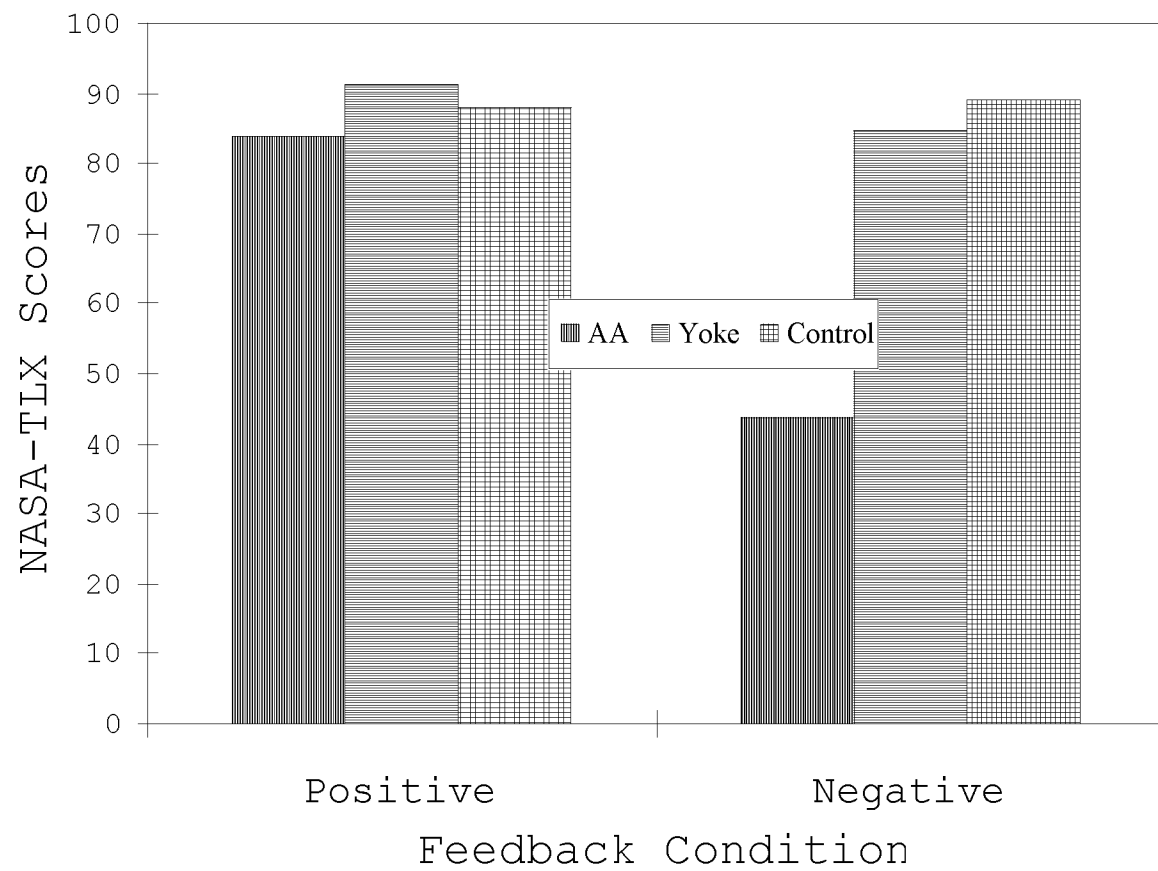


Figure 11. NASA-TLX Scores

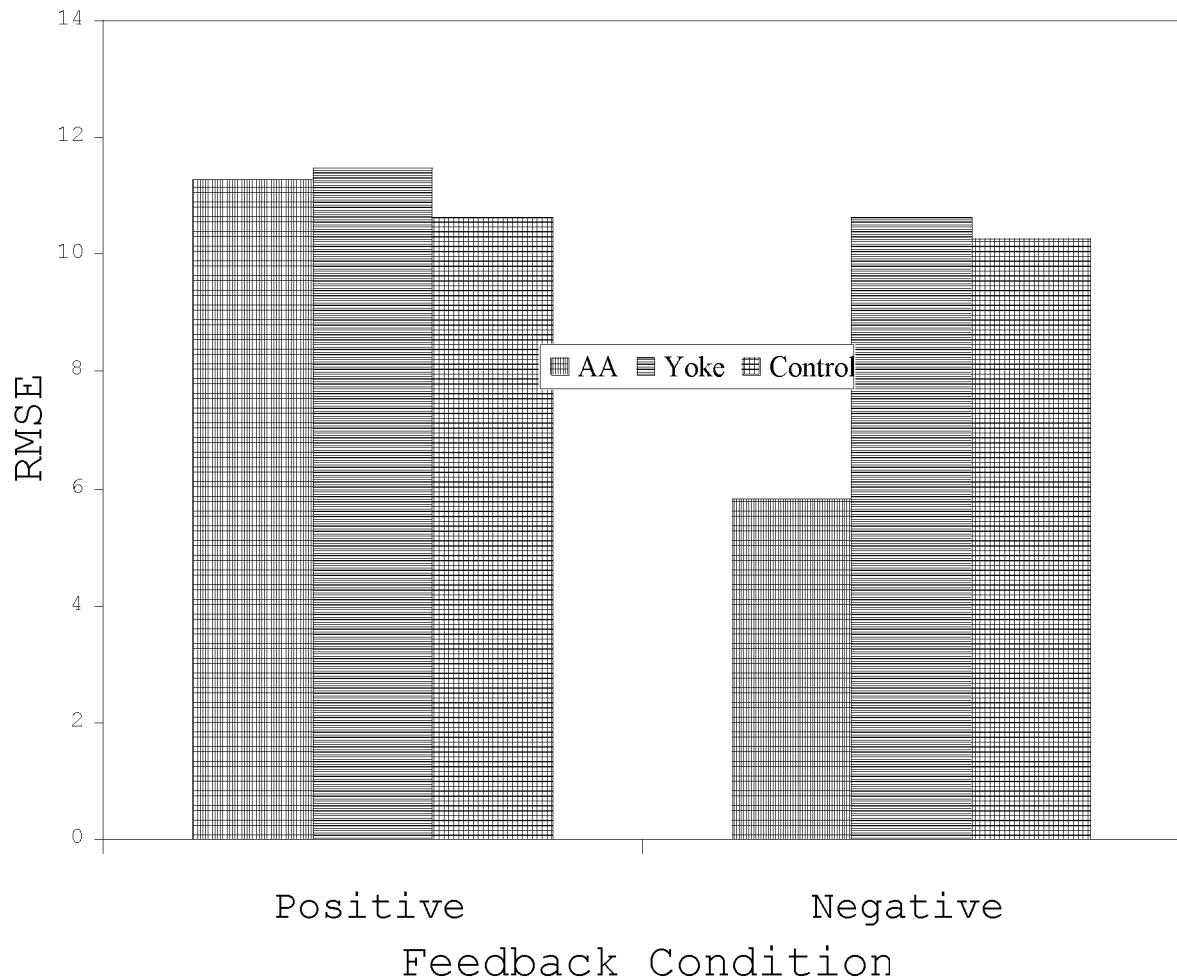


Figure 12. Tracking Performance (RMSE)

### Pilot Preference of Cycles of Adaptive Automation

Research was conducted that examined pilot preference for automation mode schedules / cycles for adaptive task allocation. Nine participants performed a tracking task and an auditory, oddball task for three trials consisting of a 15-, 30-, and 60-sec cycle durations. ERPs were gathered to infrequent, high tones presented in an auditory oddball task. The results showed that tracking performance was significantly better under the 15-sec duration, but participants rated workload significantly higher under this condition. These results were interpreted in terms of a micro-tradeoff; that is, participants did better under the 15-sec condition at the expense of working harder. The conclusion was supported by the ERP

results. An examination of the EEG gathered five seconds after each task allocation revealed that P300 latency was found to be considerably longer and the amplitude considerably smaller under the 15-sec cycle duration than under either the 30- or 60-sec cycle conditions. Therefore, these results suggest that short periods of manual reallocation may prove beneficial to performance and moderating workload demands. However, such benefits are tempered by increased “return-to-manual deficits” (Wiener & Nagel, 1988). Moreover, they support the use of ERPs metrics of workload in the design and implementation of adaptive automation technology. Note that the question of adaptive automation does not hinge on its conceptual underpinnings. Inherently, it makes sense to transform the operator's task at times when the operator's mental state is less than optimal. However, this is not to say that adaptive automation provides utility that supersedes the difficulties that we, as researchers, designers, and practitioners, may face with the implementation of this type of technology. Such studies, as those discussed previously, demonstrate that schedules of static automation can also have positive effects on performance and workload. Therefore, it is of theoretical and practical interest to determine what benefits, if any, that adaptive automation provides beyond that of static automation that cycles between automation modes based upon scripted automation schedules.

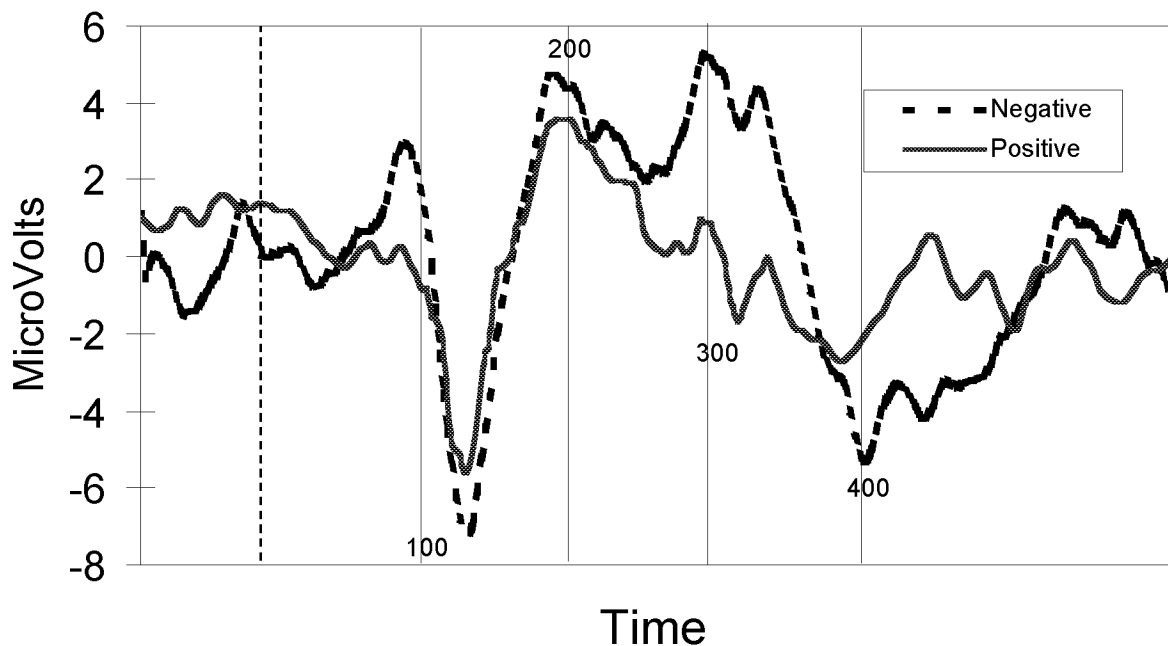


Figure 13. Event-Related Potential for Adaptive Automation Group

## Secondary Task as Metric for Adaptive Task Allocation

Our research on pilot preference for automation cycles demonstrated that adaptive task allocation, although has been shown to improve pilot performance and lower workload, could increase return-to-manual deficits and lower situation awareness, as shown by the ERP (see Figure 14), for brief periods. The word, “automation surprises” is often used to describe this concern. The question is, “What is the automation doing now?” as pilots attempt to match their mental model with the operation model of the

automation. Therefore, as Woods (1996) noted, adaptive automation represents “apparent simplicity, real complexity” and it is important to consider how such technology may affect the human-automation interaction landscape.

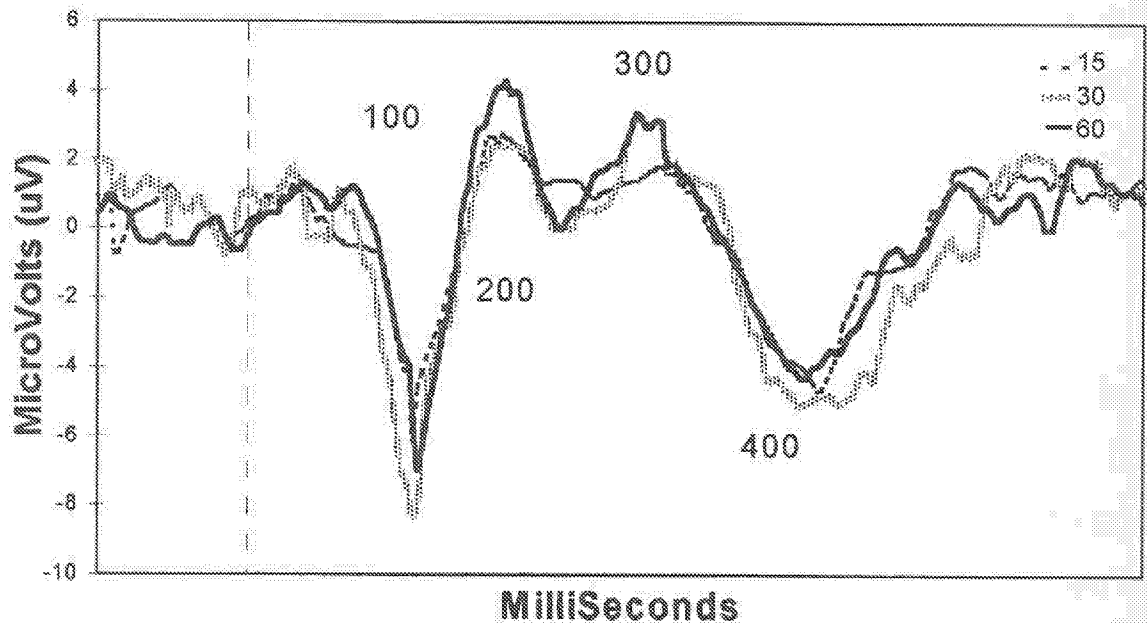


Figure 14. Event-Related Potential after Manual Reversion in Automation Mode

The problem also came out in another series of studies that were conducted to examine other ways that adaptive automation could be implemented. Secondary tasks (e.g., ATC communications) are often used as a measure of workload. The idea is that, if the primary task of flying the aircraft is high in workload, few or no cognitive resources would be available to perform secondary tasks. To examine the use of secondary task measures, various performance assessment methods used to initiate mode transfers between manual control and automation for adaptive task reallocation were tested. Participants monitored two secondary tasks for critical events while actively controlling a process in a fictional system. One of the secondary monitoring tasks could be automated whenever operators' performance was below acceptable levels. Automation of the secondary task and transfer of the secondary task back to manual control were either human- or machine-initiated. Human-initiated transfers were based on the operator's assessment of the current task demands while machine-initiated transfers were based on the operators' performance.

In the experiment, human-initiated transfers were compared to machine-initiated transfers that were based on either primary task performance or a combination of primary and secondary task performance (joint assessment). Moreover, each assessment method was tested given machine-initiated transfers to automation only and machine-initiated transfers to both automation and manual control. Altogether, there were five switching methods tested: completely human-initiated, machine-initiated transfers to automation only based on primary or joint assessment, and machine-initiated transfers to both automation and manual control based on primary or joint assessment. The five switching methods produce similar performance on the primary task measures, but there were differences among the secondary task measures (see Table 7). Machine-initiated transfers to automation coupled with human-initiated returns to manual control and joint performance assessment produced the best system performance, but these gains depended on a high reliance on automation. In addition, there was a higher

proportion of mode errors (i.e., accidental responses while in automation) given machine-initiated transfers to automation, particularly given machine-initiated transfers to both automation and manual control (see Table 7).

Table 7. Mean Performance on the Secondary Task Measures

Switching Method	Number of Trials in Automation		Hit-to-Signal Ratio		Reaction Time		Proportion of Responses Automation	
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>
Switching Method								
Automation	52.93	5.68	.925	.011			.061	.012
Both	30.08	3.41	.869	.021	No Effect		.094	.012
<u>F</u> (1,16)	45.94 ( <u>p</u> < .001)		9.80 ( <u>p</u> = .006)				10.27 ( <u>p</u> = .006)	
Task Assessment								
Joint	55.11	5.75	.937	.006	1452.55	114.13		
Primary	27.90	3.61	.857	.024	1531.25	112.32	No Effect	
<u>F</u> (1,16)	46.78 ( <u>p</u> < .001)		16.57 ( <u>p</u> = .001)		5.04 ( <u>p</u> = .039)			

The evidence indicated that machine-initiated transfers to automation with a human-initiated return to manual control produced better performance on the secondary task measures relative to machine-initiated transitions to both automation and manual control. In addition, a machine-initiated transition that considered both primary and secondary task performances yielded better operator performance on the secondary task measures and higher adjusted points relative to transition based on primary task performance alone. These gains tended to results from greater reliance on automation, though. Finally, despite the higher reliance on automation for machine-initiated transfers to automation only, this switching method produced a significantly lower proportion of mode errors compared to machine-initiated transitions in both directions.

## Neural Network Approaches to Adaptive Automation

Research has recently begun research on the use of neural nets to make automation task allocation decisions. Although the research is still in its infancy, one study has been completed that was a first look at these techniques. The study examined the applicability of information theoretic learning to develop new brain computer interfaces. It compared several features to detect the presence of event-related desynchronization (ERD) and synchronization (ERS) and developed several classifiers for temporal pattern recognition in the EEG record.

Data was collected from 6 sessions with three subjects (g3, g7, and i2) that performed at three levels of performance (very good, moderate, and poor) during pre-testing. Four electrodes were placed 2.5 cm anterior and posterior to C3 and C4 to obtain two bipolar EEG channels over the left and right motor areas. The signal was sampled at 128 Hz, with an anti-aliasing filter at 30 Hz. The trials were visually inspected for artifacts and these trials removed. Figure 15 show that the alpha (9-13 Hz) and beta band (20-24 Hz) contained the information for the ERD and ERS. A 5<sup>th</sup> order Butterworth IIR filter was used and was integrated for 1 second to obtain signal power. Downsampling to 8 Hz was done next.

Therefore, using the a priori knowledge about the task, it was possible to quantify the EEG activity with a 4-feature vector (alpha and beta band powers at the left and right motor areas).

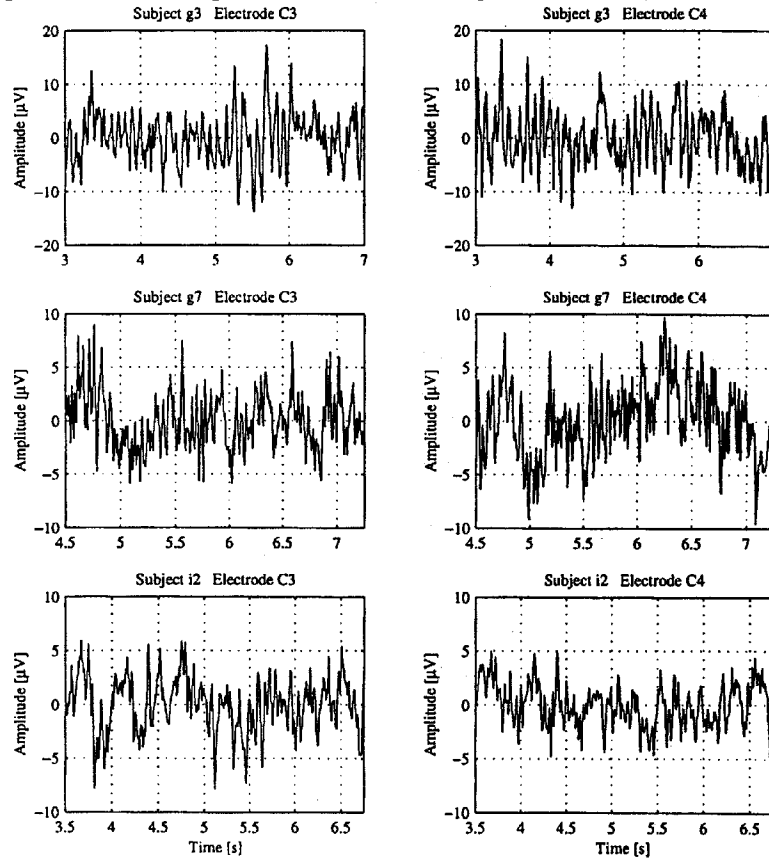


Figure 15. ERD and ERS

The median Bhattacharyya distance (MBD) measure was used as a measure of separation between left and right hand movements, and therefore as a measure of performance. Then, three alternative features besides alpha and beta band powers were assessed and compared: (1) autoregressive (AR) model using the RLS (recursive least square) algorithm. The information is contained in the coefficients, so the averaging was performed over 16 time steps and the feature set was 12 dimensions; (2) Hjorth parameters which involves the analysis of activity, mobility, and complexity as three time domain descriptors of EEG activity. Activity is simply the variance of the signal, the mobility is the ratio of the variance of the first derivative of the signal over the variance of the signal, while complexity is the ratio of the mobilities of the first derivative and of the signal itself. These were averaged over 16 time steps, providing a 6 dimension vector; (3) and finally, a Principle Component Analyzer (PCA) with the Sanger rule which is an optimal filter bank where the square of the outputs are the eigenvalues of the time autocorrelation matrix. Figure 16 shows that the MDB was different across the three subjects. Subject g3 provides the largest separation and hence will yield the best classification results whereas i2 yielded the poorest.

One of the difficulties of temporal pattern recognition faced in training dynamic neural networks is what is the desired response over time. A comparison was made of conventional classifiers: a linear discriminant based on Fisher's method, a perceptron, a time delay neural network (TDNN), a gamma network, and a TDNN trained with dynamic targets (DT). Table 8 shows the different error rates for different classifiers (%).

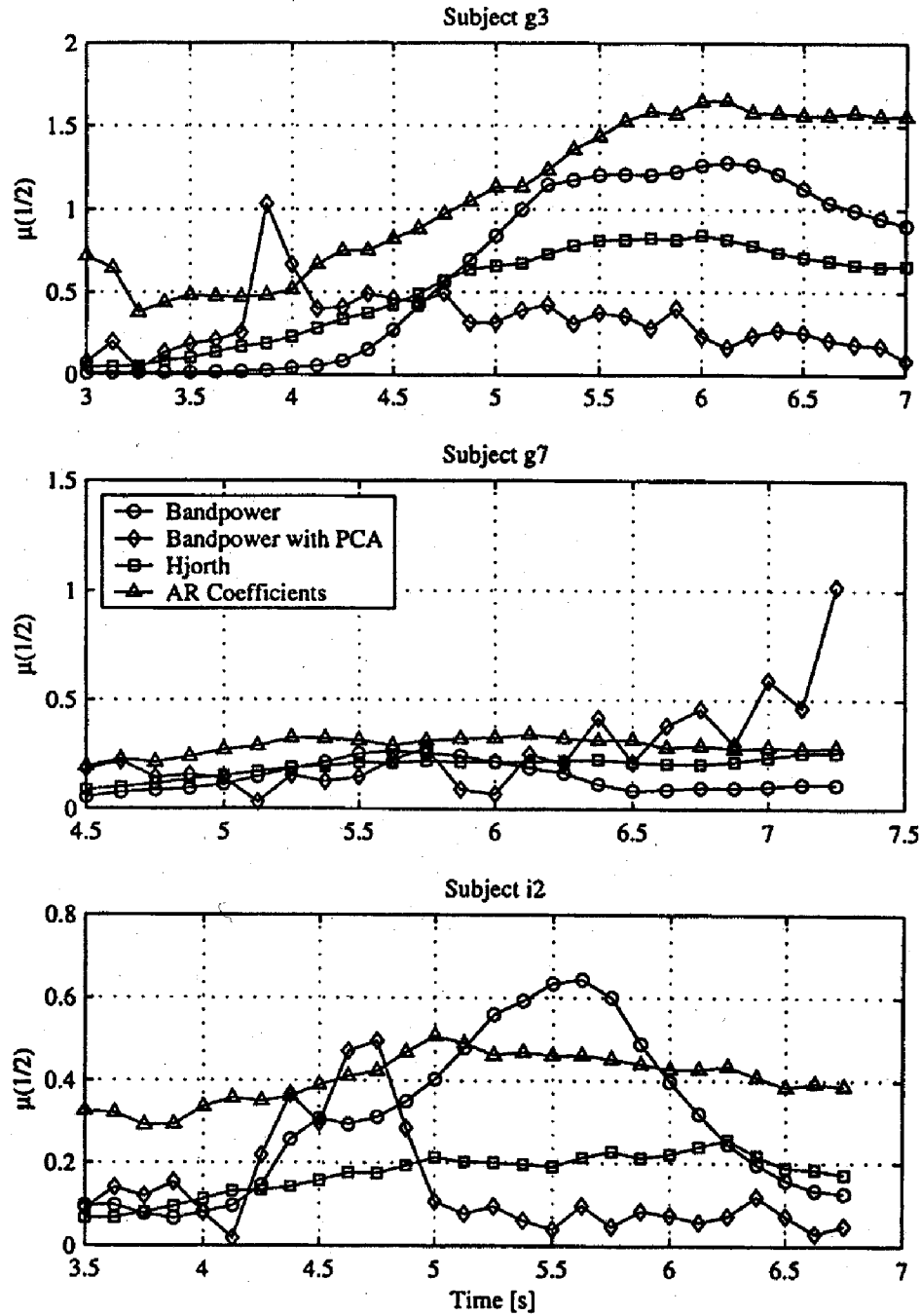


Figure 16. MDB across windows for the 3 subject and different preprocessors

The table shows that the neural network topology only affects slightly the general trends of accuracy. If a subject was able to create ERD/ERS (as subject g3) then any of the classifiers work reasonably well. For simplicity, however, the linear Fisher discriminator was best. However, for top performance, a neural



network would have to be trained over time. The outcome of the study was that participants could interface and affect control inputs through these physiological signals created at the motor cortex. The significance of these results is that these theoretic information-learning algorithms could be employed to control many different types of human-machine interfaces including aircraft.

Table 8. Error Rates for Different Classifiers (%)

subject	Fisher	MLP	TDNN	gamma	DT
g3	7	6.8	7.1	7.7	4.8
g7	32	34	30.8	34	32.4
i2	22	20	21	20.6	20.8

The work on neural networks for ERD/ERS has not been continued because adaptive automation will probably require other physiological inputs than motor cortex responses. However, the research was productive in helping to establish the efficacy of these classifiers for neural nets that may potentially drive the operation of a real-time adaptively automated system whether based on psychophysiology, performance, operator performance models, etc.

## Conclusions of Research on Adaptive Automation

The system that has been developed and research at NASA Langley Research Center has made valuable contributions to the field of automation design. Specifically, the research has made significant progress in understanding the use of psychophysiology to adaptively change task allocation so as to optimize human-automation interaction. Although these works provide the most comprehensive body of research in the use of psychophysiology for adaptive automation, there is a significant amount of work that is still needed. Adaptive automation has the real potential to significant change how pilots interact with automation in the cockpit. Furthermore, this area of automation has branched into other areas of aviation including maintenance and ATC work. Therefore, it is important that fundamental research be conducted to ensure that many of the same problems that exist today with current automation do not merely transform into similar or worse problems with adaptive automation. If adaptive automation is to have a place in aviation, clearly there has to a positive cost-benefit demonstrated that shows that this form of automation minimizes issues of cognitive overhead, complacency, under-reliance, clumsy workload, automation surprises, mode unawareness, loss of situation awareness, etc.

That is the general conclusion. However, specific to our work and the biocybernetic system, it is unknown what role psychophysiology may play in the development of adaptive automation. Although other programs (e.g., Pilot Associate's Program) and eminent researchers in the area have endorsed the use of psychophysiology (e.g., Bryne & Parasuraman, 1996), the use of psychophysiology as a "trigger" to invoke adaptive aiding or adaptive allocation may not be feasible currently. Despite the need for future work, the research thus far has been very successful in demonstrating that a system could couple the level of automation with the level of operator engagement thereby facilitating performance and lower workload with a psychophysiology based adaptive interface. Also, it should be noted that, regardless of whether psychophysiology may serve in the "predictive" role or dynamic changes to the automation landscape in real-time, Byrne and Parasuraman (1996) noted that the approach could be used for evaluating other potential parameters in an adaptive automation framework (the "developmental" role). In fact, our research on secondary task measures has shown this to be case. In the section of "current

research focus” of the physiological factors element, this approach is discussed in regards to the research that is presently being conducted that uses psychophysiology in conjunction with other approaches to form a global means of keeping the pilot “in-the-loop” through the dynamic regulation and modulation of task engagement and automation mode.

## INDIVIDUAL DIFFERENCE VARIATE RESEARCH

Although adaptive automation has remained the fundamental focus of our research, the basic foundation of the work relies on the concept that there are hazardous states of awareness that provide the potential for errors to take place in aviation. Therefore, another line of research has been directed towards the understanding of what may play a role in inducing the states. The etiologies can be complex and can take a number of different paths. One of these paths that we have addressed in a small fashion is the notion of individual difference variates as potentials for the onset of HSAs. During FY99-01, we conducted two different studies that have looked at the issue: (1) self-efficacy and (2) boredom proneness, cognitive failure, and complacency potential and the role of these personality predisposition in contributing to a particularly deleterious and timely hazardous state of awareness – automation-induced complacency.

### **Self-Efficacy and Automation-Induced Complacency**

Crew “complacency” has often been implicated as a contributing factor in aviation accidents. Complacency has been defined as “self-satisfaction which may result in non-vigilance based on an unjustified assumption of satisfactory system state” (Billings et al., 1976). The term has become more prominent with the increase in automation technology in modern cockpits. As a consequence, there has also been an increase in research to understand the nature of complacency and to identify countermeasures to its onset. Parasuraman, Molloy, and Singh (1993) noted that complacency arises from overconfidence in automation reliability. They found that operators missed “automation failures” when the automated system was highly reliable. Riley (1996) reported that an operator’s decision to rely on automation may actually depend on a complex relationship between level of trust in the system, self-confidence, and other factors. Lee and Moray (1994) also found that trust in automation and self-confidence can influence decisions to use or not to use automation, but that there were large individual differences. The idea of individual differences was examined recently by Singh, Molloy, & Parasuraman (1993a). They reported a modest relationship between individual differences in complacency potential and energetic-arousal with automation-related monitoring inefficiency.

Research was conducted in 2000 to further explore the effects of individual differences in automation use. Specifically, we examined the generalizability of self-efficacy in monitoring performance and its relationship to automation-induced complacency. Self-efficacy refers to expectations that people hold about their abilities to accomplish certain tasks. Bandura (1986) argued that decisions to undertake particular tasks depend upon whether or not they perceive themselves efficacious in performing those tasks. The stronger the operator’s self-efficacy, the longer they will persist and exert effort to accomplish the task (Garland et al., 1988). Studies have shown that people with higher self-efficacy perform better compared to people with lower self-efficacy. However, in the aviation context, conditions do arise in which self-efficacy and the concomitant overconfidence in one’s ability can impair performance. As an example, a pilot not off-loading tasks to automation during high workload situations because of overconfidence in managing flight tasks. Therefore, we were interested in examining the effects of self-efficacy on automation use and complacency under high and low workload conditions.

Thirty participants performed a 30-min vigilance task that required responses to critical events. There were 30 critical events presented during the vigil and the event rate was 30 per minute. Afterwards, each participant was asked to complete both general and task-specific self-efficacy questionnaires

(Schwarzer & Jerusalem, 1995) as well as the Complacency Potential Rating Scale (Singh, Molloy, & Parasuraman, 1993b). There was no statistical difference between the participants in task performance. These participants were then assigned to two experimental groups based on a median split of the self-efficacy questionnaires. All participants returned after one week and performed a system monitoring, resource management, and tracking task from the Multiple Attribute Battery (Comstock & Arnegard, 1992) under high reliable and low reliable conditions (see Parasuraman, Molloy, & Singh, 1993 for description). The difficulty of the tasks varied during each task run, and participants had an option to off-load the tracking task to the automation.

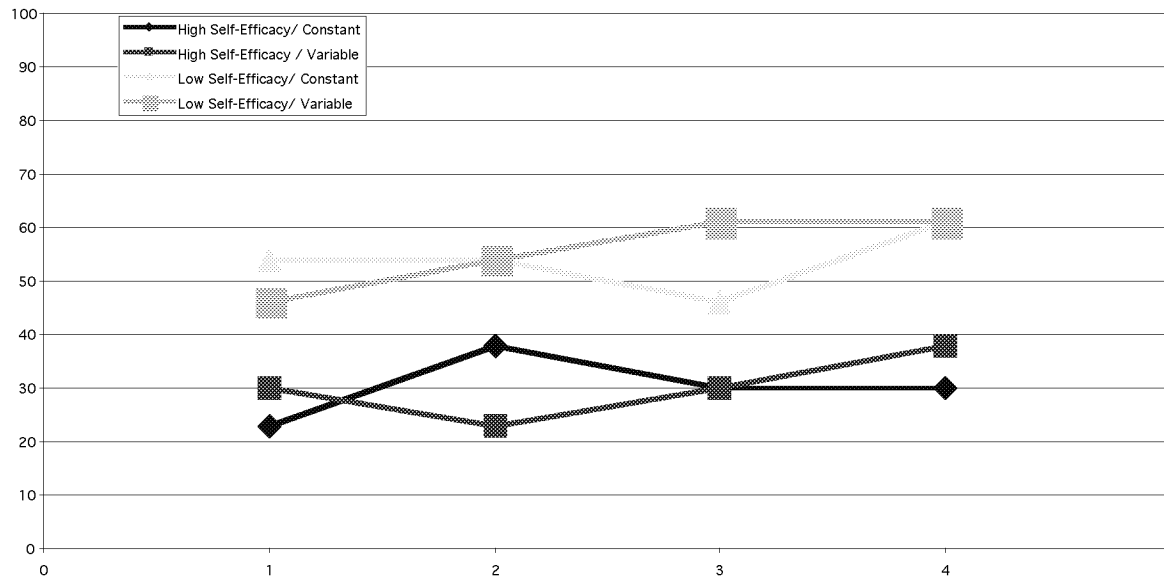


Figure 17. % Hits (y axis) X Session (4) for High Workload Condition

Our findings were that participants rated high in self-efficacy performed significantly better and had lower complacency scores. However, under conditions of high workload, these participants failed to off-load the tracking task and performed significantly worse (Figure 17) and rated workload higher (Figure 18) than participants rated lower in self-efficacy. These results suggest that self-efficacy is an important component of whether an operator will succumb to automation-induced complacency. However, self-efficacy may serve as a double-edged sword in producing overconfidence in one's ability that may limit other strategies, such as task off-loading, for managing workload.

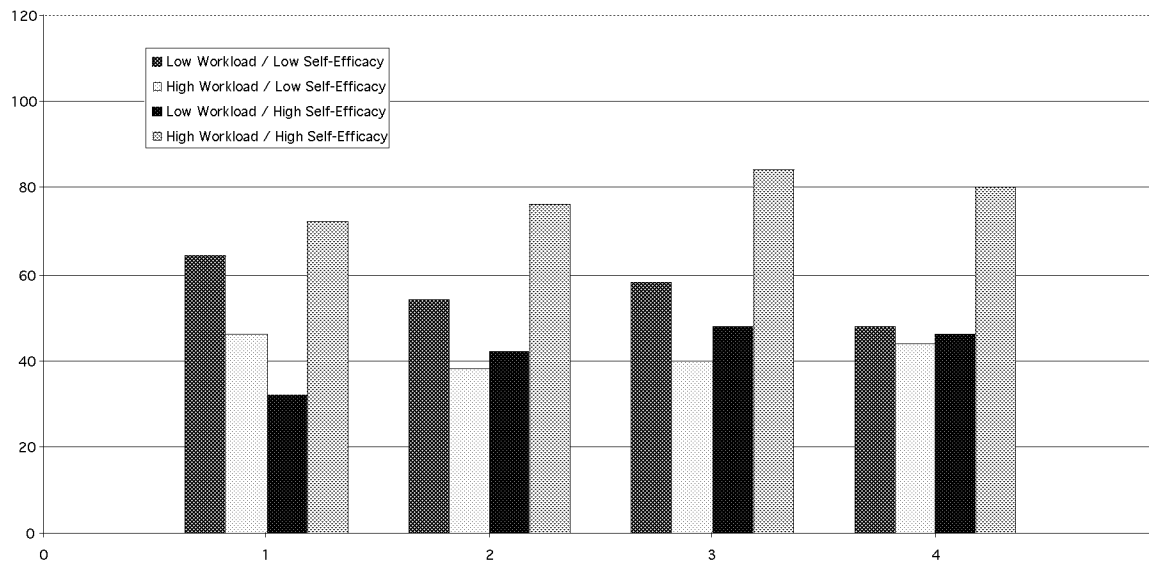


Figure 18. NASA-TLX Scores

## Boredom Proneness, Cognitive Failure, and Workload

Mental workload refers to the amount of processing capacity that is expended during task performance (Eggemeier, 1988). Riley (1996) noted that although workload was a necessary aspect of automation-induced complacency, little workload-related research exists. Parasuraman and his colleagues (1993), found the low workload level of a single task condition, consisting of only a system-monitoring task, was not sufficient to induce complacency. They reasoned that in a single-task environment a state of boredom would be experienced by the subjects, due to the low workload level involved in the task. The detection rates, however, for both reliability groups in this condition were extremely high (near 100%). Therefore, they concluded that the lack of complacency experienced by participants in the single-task condition suggested that complacency and boredom are two distinct concepts.

In contrast, several studies have linked boredom, especially the propensity to become bored, to high amounts of workload. Sawin and Scerbo (1994, 1995) in their use of vigilance tasks report that boredom often has a high workload aspect associated with it. The information-processing demands or workload experienced by participants performing a vigilance task were once thought to be minimal. Fulop and Scerbo (1991), however, have recently demonstrated that participants find vigilance tasks to be stressful and other researchers have found them to be demanding due to the high workload involved in remaining vigilant (Deaton & Parasuraman, 1993; Galinsky, Dember, & Warm, 1989).

**Boredom Proneness.** Farmer and Sundberg (1986) isolated a single measurable trait, *boredom proneness* (BP), which they report as highly related to a person's tendency to become bored. They developed a 28-item scale, the Boredom Proneness Scale (BPS; Farmer & Sundberg, 1986), to measure this trait. Stark and Scerbo (1998) found significant correlations between workload, complacency potential, and boredom proneness, by examining their effects on task performance using the Multi-Attribute Task Battery (MAT; Comstock & Arnegard, 1992). Their study supports the view that the psychological state of boredom may be a factor that induces complacency. The results of Parasuraman et al. (1993) thus need to be considered cautiously since they reported no workload or boredom data to

support their claim that their single task represented an underloaded task condition which caused boredom and, therefore, that boredom and complacency are unrelated. A considerable amount of evidence points to high workload being associated with boredom components while performing supervisory control and vigilance tasks (Becker, Warm, Dember, & Hancock, 1991; Dittmar, Warm, Dember, & Ricks, 1993; Prinzel & Freeman, 1997; Scerbo, Greenwald, & Sawin, 1993). In addition, Pope and Bogart (1992) reported that ASRS reports contain descriptions of crews becoming “complacent” due to succumbing to “boredom” and “experiences of diminishing attention, compromised vigilance, and lapsing attention, frequently not associated with fatigue” (p. 449). Therefore, automation-induced complacency is composed of a number of dimensions including trust, boredom proneness, complacency potential, self-confidence, skill-level, workload management ability, and experience, to name a few. All of these dimensions are or can be influenced by the individual differences of each human operator. For example, Riley (1989) stated that trust is a multidimensional construct that has both cognitive and emotive qualities that can be influenced by individual differences.

**Cognitive Failure.** Grubb, Miller, Nelson, Warm, and Dember (1994) examined one such personality dimension, “cognitive failure” and its relation to perceived workload in vigilance tasks, as measured by the NASA-TLX. They reported that operators high in cognitive failure (HCF) tend to be more absent-minded, forgetful, error-prone, and less able to allocate mental resources to perform monitoring tasks than those classified as low in cognitive failure (LCF; Broadbent, Cooper, Fitzgerald, & Parkes, 1982). Interestingly, Grubb et al. found, HCF and LCF participants performed equally well on vigilance tasks but the workload scores of the HCF were significantly higher than their LCF peers; thus, these participants performed as well as LCF participants but did so at a higher cost in resource expenditure. The HCF individuals, therefore may exhibit complacent behaviors, due to their resources being largely depleted, when faced with continuing a task. This prevalence towards cognitive failure may be another factor related to a person’s becoming complacent while monitoring automation.

The individual differences described above suggest that automation-induced complacency may represent a complex dynamic of many psychological constructs. As Singh et al. (1993) describe, “...the psychological dimensions of complacency and its relation to characteristics of automation are only beginning to be understood...”and that other individual and social factors may also play a role. Therefore, a need remains to examine other psychological antecedents that may contribute to automation-induced complacency.

We conducted a study to examine automation-induced complacency in relation to the personal dimensions of complacency potential, boredom proneness, and cognitive failure. All of these dimensions are hypothesized to have an affect on whether an individual will experience complacency within a multi-task environment. 40 participants performed tasks on the MAT under conditions of constant or variable reliability that have been found to induce complacency in previous studies (e.g., Parasuraman et al., 1993). The three individual difference measures, Complacency-Potential Rating Scale (CPRS; Singh et al., 1993), Cognitive Failure Questionnaire (CFQ; Broadbent et al., 1982), and the Boredom Proneness Scale (BPS; Farmer and Sundberg, 1986) were gathered to measure these traits in each participant. The NASA-TLX (task-load index; Hart & Staveland, 1988) and the Task-related Boredom Scale (TBS; Scerbo et al., 1994) was also used to assess the total subjective workload and total perceived boredom experienced by each participant.

Participants assigned to the high complacency potential group ( $M = 25.87$ ) scored higher on the boredom proneness scale than participants assigned to the low complacency potential group ( $M = 14.85$ ),  $F(1,39) = 67.31$ ,  $p < .0001$ . A significant interaction for Complacency Potential (CP) x Reliability Condition (RC) was found,  $F(1, 39) = 4.58$ ,  $p < .05$ . Individuals found to be high in complacency potential were more likely to exhibit performance decrement illustrative of automation-induced complacency. Also, participants assigned to the high complacency potential group rated overall mental workload on the NASA-TLX ( $M = 57.05$ ) to be significantly higher than participants in the low complacency potential group ( $M = 46.67$ ),  $F(1,39) = 6.82$ ,  $p < .01$ . In addition, participants in the variable reliability condition ( $M = 28.94$ ) performed significantly worse overall on the tracking task than

participants in the constant reliability condition ( $M = 17.20$ ),  $F(1, 39) = 28.12$ ,  $p < .0001$ . Furthermore, participants assigned to the high complacency potential group ( $M = 30.15$ ) also had higher tracking RMSE overall than participants in the low complacency potential group ( $M = 15.98$ ),  $F(1, 39) = 40.89$ ,  $p < .0001$ . There was also a significant interaction between Complacency Potential and Reliability Condition,  $F(1, 39) = 8.63$ ,  $p < .005$ . Finally, there was an interaction of CP x RC for A', however has strong implications for the study's hypotheses,  $F(1, 39) = 11.49$ ,  $p < .001$ . Participants across all groups and conditions performed comparably with the exception of the high CP x constant RC who did significantly poorer.

**Implications of results for complacency.** What do these results mean? As predicted, the complacency potential rating scale successfully discriminated those individuals who were more likely to "trust" and "overrely" on the automation during conditions of automation reliability that produce such behaviors. That effect was not surprising as the scale had been validated in previous studies (Singh, Molloy, and Parasuraman, 1993). However, we were interested in what was it about complacency potential that produced this hazardous state of awareness and our previous research suggested that it was actually an interaction of operator strategy (e.g., trust) and certain other hazardous states of awareness (e.g., boredom). These effects were demonstrated especially with regards to boredom proneness and cognitive failure.

Fahlgreen and Hagdahl (1990) asked over 1,000 pilots to provide definitions of complacency and, although there were varying definitions, the best one was "A mental state where a pilot acts, unaware of actual danger or deficiencies. He still has the capacity to act in a competent way – but for some reason, or another, this capacity is not activated. He has lost his guard without knowing it." The results of the present study have implications for the study of complacency. Wiener (1981) stated that complacency results when pilots are 'FDH: Fat, Dumb, and Happy.' It is familiarity, experience, and trust that fosters this psychological condition that influences our expectations and expectations control our perceptions. The definition provided by Fahlgreen and Hagdahl suggests that complacency is almost unconscious. Unsafe acts and deliberate violations of Standard Operating Procedures (SOP) and regulations is not complacency. Rather, as Jensen (1995) noted, "*Unconsciously* following rules, SOPs, automatic systems, management edicts, without question and sound justification is complacency" (p. 242). Complacency has been linked to the vigilance problem in which self-satisfaction sets in and acts are performed without thinking – automatic processing (Shneider & Shiffrin, 1990) – and the risk of missing something important increases (i.e., errors of omission). Pilots blame the majority of their errors of omission on complacency and, therefore, the construct has definitive shared linkages with other psychological constructs including those of boredom and mental workload.

Such a perspective shared by pilots, air traffic controllers and other operators as well as some, but not all (e.g., Parasuraman et al., 1993) researchers, were supported by the results reported here. Boredom proneness and task-related boredom bore significant correlations with complacency potential and task performance. Parasuraman et al. (1993) noted that complacency and boredom were not related because, in their study, the low workload condition did not produce automation-induced complacency. Boredom, however, is actually a heightened state of arousal and a frustration because of a need to do something. Scerbo, Greenwald, and Sawin (1993) and Sawin and Scerbo (1995) reported that NASA-TLX Frustration subscales were significantly correlated with self-reports of boredom. Boredom is a very distressing condition for humans and, given the choice, most humans would prefer to be frightened than bored (Jensen, 1995). Boredom often occurs when a task is highly mastered and routine and no longer presents a challenge. Complacency can set in because of a need to seek novel stimulus and mental stimulation. Myers and Miller (1954) and Isaac (1962) reported that animals and humans have an exploration drive that is instigated by boredom. Therefore, when we move away from our state of motivational equilibrium, a natural inclination is to balance congruity (Deci, 1975). Mackworth (1970) reported that repetition and monotony could produce habituation in behavioral and physiological response. There are neurophysiological theories that implicate neural inhibition, reduction in arousal levels, and changes in characteristics of evoked potentials to explain decrements found in vigilance

behaviors and resultant habituation of responses that lead to errors of omission. A number of researchers (Sharpless & Jasper, 1956; Yerkes & Dodson, 1908) have noted the almost “daydream” state produced by habituation. The results of the present study suggest further that certain people may be more susceptible to the deleterious effects that boredom can produce in task performance --- namely, complacency.

Related to boredom proneness, cognitive failure was also found to provide diagnosticity in differentiating those participants who would exhibit the behaviors of automation-induced complacency. The participants rated high in cognitive failure tended to be more absent-minded, forgetful, error-prone, and less able to allocate mental resources to perform monitoring tasks. Under the multitask situation, the mental workload associated with performing the task impacted the availability of cognitive resources to perform the tasks and may have developed a strategy for trusting the automation. Parasuraman et al. (1993) stated that the allocation of cognitive resources was not indicated in producing automation-induced complacency behaviors. But, taking a Multiple Resource Theory (Wickens, 1985) perspective, the increase in taskload in a multiple task situation has theoretical effects on the availability of cognitive resources to perform the task. The Cognitive Failure scale differentiates those who are less able to allocate resources to manage workload demands and, therefore, may have to revert to other strategies such as trusting the automation. Such a conclusion is supported by the increase self-reports of mental workload, measured by the NASA-TLX, for participants rated high in Cognitive Failure.

## TRAINING FOR HAZARDOUS STATES OF AWARENESS RESEARCH

### **Psychophysiological Self-Regulation Training**

Adaptive automation is still in its conceptual phase and a number of research issues still need to be addressed before widespread acceptance will be possible. Woods (1996), for example, noted that automation represents “apparent simplicity, real complexity” referring to the idea that automation transforms the nature of pilot-automation interaction. New forms of automation may bring with them new problems. Adaptive automation may not be an exception to such an observation. Rudisill (1994) found that pilots tended to be positive about technological innovations, but still had concerns about advanced forms of automation. They noted that advanced automation kept them “out-of-the-loop” and that pilots constantly needed to monitor what the automation was doing as well as increasing their workload and decreasing cockpit management and flight crew communication. Pilots reported that what they wanted were new approaches to training that would ameliorate some of these problems associated with advanced automation.

Research at NASA Langley Research Center has developed a training method that may complement the use of adaptive automation. The research was stimulated because of past research in our laboratory in which we found increased workload and had increased return-to-manual deficits in which task performance suffered significantly just after a task allocation under adaptive automation conditions. Therefore, we posited that adaptive task allocation would be best reserved at the endpoints of the task engagement continuum and that other techniques should be used in conjunction with adaptive automation to help minimize the onset of hazardous states of awareness (Pope & Bogart, 1992) and keep the pilot “in-the-loop.” One training technique that may be employed is psychophysiological self-regulation.

Psychophysiological self-regulation refers to the ability of a person to control affective and cognitive states based on autonomic (ANS) and central nervous system (CNS) functioning. The techniques use physiological markers of these states and provide feedback so that the person learns these associations and how to modulate their occurrence. The training technique we present here uses neurofeedback from the electroencephalogram record to help control the onset of hazardous states of awareness.

Currently, there has not been much research conducted on the use of physiological self-regulation for performance enhancement (see Norris & Currier, 1999 for review). With regards to aviation, there

has been virtually no research examining the efficacy of self-regulation for improving pilot performance. One of the few studies that have been conducted was reported by Kellar et al. (1993). They found that self-regulation training, termed “Autogenic-Feedback Training (AFT)”, may be an effective countermeasure to stress-related performance decrements. Additionally, these authors reported that AFT improved crew coordination and communication and, therefore, may serve as a valuable adjunct to CRM training. Although Kellar et al.’s (1993) research demonstrated the value of physiological self-regulation for controlling stress-related responses to emergency conditions, stress represents only one of the hazardous states of awareness that pilots may encounter during flight. Other states include boredom, inattention, complacency, fatigue, etc. that may play an equal or greater role in contributing to incidents and accidents in aviation.

Research has shown promise for the closed-loop system to serve in both regulatory and developmental roles for the use of psychophysiology in adaptive automated systems (Byrne & Parasuraman, 1996). To date, however, our research had focused on the examination of system parameters for the real-time task allocation of automation modes (i.e., manual; automatic). Furthermore, the task mode alone was responsible for determining task allocation sequencing; that is, what automation mode the task was in determined the engagement level of the participants which therein determined subsequent task modes. However, research in biofeedback and self-regulation has demonstrated the capabilities that people have to control their own engagement states. Therefore, considering the theoretical foundation that the system is based upon, it seems reasonable to explore the biofeedback potential of the system as a training tool for developing self-regulation skills for managing hazardous states of awareness.

To examine the efficacy of physiological self-regulation, participants were assigned to three experimental groups (self-regulation; false feedback; control). The self-regulation group was provided neurofeedback training that focused on learning the patterns of hazardous states of awareness and performance knowledge-of-results (KR). To guard against the chance that just providing feedback may be responsible for producing positive effects, the false feedback group was given random feedback regarding their mental engagement state and performance. The control group was provided no feedback. It was hypothesized that physiological self-regulation training would provide tools for participants to manage their cognitive resources by self-regulation of their engagement states. The expected outcomes of which would be better performance, lower reported subjective workload, and fewer automation task allocations for these participants compared to those in the false feedback and control groups.

**Method.** Eighteen participants performed tasks from the MAT and the tasks were changed based on the adaptive MAT, biocybernetic protocol described previously. There are six levels by which the system determines task allocation. Levels 1-3 reflect decreasing engagement and levels 4-6 reflect increasing engagement relative to baseline measures. Within these two categories, levels are determined based upon how variable the EEG engagement index was during baseline performance. The algorithm used to determine the level of the task the participant would be placed is as follows: A level 3 allocation would be assigned if the index was between 0 and  $-0.5$  standard deviations (SD) below the baseline mean; level 2 would be assigned for an index between  $-0.5$  to  $-1.00$ ; level 1 would be assigned if the index was below  $-1.00$ . For indexes above the baseline mean, level 4 would be assigned if the index was between 0 and  $+0.5$  SD above the mean; level 5 between  $+0.5$  and  $+1.00$ ; and level 6 above  $+1.00$ .

All participants were instructed that the system measured six different engagement levels, and that a high difficulty, manual task allocation would occur if the engagement level went to Level 1 (low task engagement) or automatic task allocation if it went to Level 6 (high task engagement). When the engagement level was between Levels 2 and 5, the tracking task was in the manual, low difficulty task mode. If the index indicated that the participant's arousal level was 1 SD above baseline (level 6), the task was switched from the manual task condition to the automatic task condition. If the index indicated that arousal was 1 SD below baseline (level 1), the task was switched from either the automatic or low difficulty, manual task condition to the high difficulty, manual task condition.



There were three separate experimental groups for this study (self-regulation, false feedback, control). Participants in the self-regulation group were provided biofeedback regarding their task engagement level while they participated in two 30-minute training sessions. During the first training session, feedback on engagement level was provided in the right-hand corner of the tracking window (e.g., “Level One”) during the training sessions, and they were encouraged to try and maintain Level 3 or 4 engagement levels. Furthermore, these participants were provided knowledge-of-results (KR) feedback on their performance as to their performance (root-mean-squared-error; RMSE) during the experimental session. The feedback was provided in the lower left-hand box of the tracking window. During the second training session, participants were cued by a computer-generated tone to estimate what their engagement level was at particular times during the session (i.e., pressing F1-F6 keys that corresponded to engagement levels 1-6). Feedback was then provided as to how close their estimation was to their actual engagement level. All participants in the self-regulation condition achieved a 70% level of correct identifications.

The false feedback group was presented with identical training protocols as in the self-regulation condition. The only exception was that these participants were provided incorrect feedback as to their task engagement level and performance. False feedback was given as +/- 1 engagement level and +/- 5 RMSE from actual task engagement and performance levels. These values were determined during pilot testing in which participants commented that these incorrect values seemed realistic to their current state and performance (i.e., the false feedback provided enough diagnosticity as to be believable). Participants in the control condition were not provided with any feedback concerning their task engagement and performance, but these participants did complete two 30-minute “no training” sessions.

**Results.** A main effect was found for tracking performance,  $F(2,15) = 82.86$ ,  $p < .0001$ . Participants in the self-regulation group performed significantly better ( $M = 2.03$   $SD = 0.28$ ) than participants in either the false feedback group ( $M = 7.77$   $SD = 0.90$ ) or control group ( $M = 6.62$   $SD = 0.87$ ). Furthermore, return-to-manual deficits were found to be higher for participants in the control condition ( $M = 15.43$   $SD = 3.98$ ) and false feedback condition ( $M = 16.89$   $SD = 4.21$ ) than participants in the self-regulation condition ( $M = 9.87$   $SD = 2.56$ ),  $F(2,15) = 10.45$ ,  $p < .05$ . Figure 19 represents tracking RMSE across each 10-minute experimental block.

An EEG difference score was calculated by subtracting the mean for each participant’s task EEG Engagement Index from the mean of his or her baseline EEG engagement index ( $EEG\ Index_{task} - EEG\ Index_{baseline}$ ). The EEG difference score was found to be significantly smaller in the self-regulation condition ( $M = 2.73$   $SD = 2.19$ ) than in either the false feedback ( $M = 14.36$   $SD = 6.61$ ) or control ( $M = 12.86$   $SD = 8.24$ ) conditions,  $F(2, 15) = 6.18$ ,  $p < .01$ . EEG engagement index values for each condition were: Self-regulation ( $M = 17.00$   $SD = 6.08$ ), false feedback ( $M = 24.60$   $SD = 3.28$ ), and control ( $M = 28.94$   $SD = 12.10$ ). No significant differences were found between the three groups’ baseline EEG engagement index ( $p > .05$ ). Figure 20 shows the EEG difference score across each 10-minute experimental block.

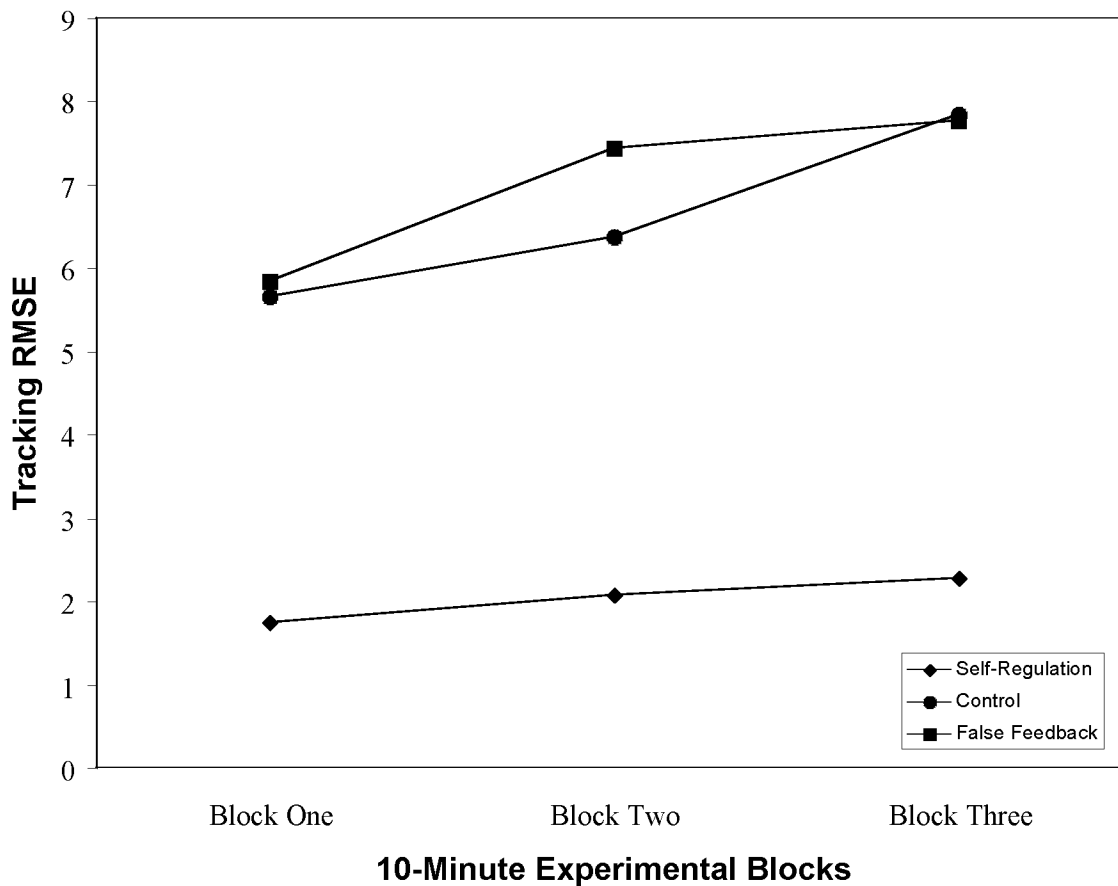


Figure 19. Tracking RMSE across 10-Minute Experimental Blocks

An ANOVA revealed that participants in the self-regulation group ( $M = 38.00$   $SD = 12.08$ ) rated workload to be significantly lower than participants in either the false feedback ( $M = 58.66$   $SD = 16.46$ ) or control groups ( $M = 66.66$   $SD = 17.28$ ),  $F(2,15) = 5.50$ ,  $p < .05$ .

An ANOVA was performed on the number of task allocations made between automation levels. The analysis was done because the intention of self-regulation training is to reduce the need to make task allocations in order to keep the operator “in-the-loop.” A main effect was found between conditions for number of total task allocations,  $F(2,15) = 7.52$ ,  $p < .01$ . There were significantly fewer task allocations made in the self-regulation condition ( $M = 19.00$   $SD = 7.79$ ) than in either the false feedback ( $M = 40.50$   $SD = 12.09$ ) or control conditions ( $M = 40.33$   $SD = 12.59$ ). An examination of Figure 21 shows that most of the task allocations made in the self-regulation condition were confined to Levels 3 and 4 which was considered optimal for task engagement and performance. Task allocations in the other two conditions were spread roughly equally across the automation levels.

Although participants in the false feedback and control groups had more task allocations to the automated and difficult, manual task conditions, these participants spent only approximately 10% and 12% of their time in either of these two task conditions (automated and difficult, manual, respectively). Participants in the self-regulation group, however, also spent approximately 11% and 9% of their time in the automated and difficult, manual task conditions, respectively. Therefore, the differences found in performance cannot be attributed solely to different task demands since each group did perform all three task conditions for equal amounts of time.

**Conclusions.** Sarter and Woods (1994) claimed that, with the presence of multiple modes of automation, flying becomes a task of orchestrating a “suite of capabilities” for different sets of circumstances. For example, Endsley & Kiris (1994) found that higher levels of autonomy remove the operator from the task at hand and can lead to poorer performance during automation failures; a problem

that may be more acute with increasing numbers of task allocations between modes. Scerbo (1996) noted that automated systems with multiple modes are difficult to learn and may increase the workload associated because the intention of system behavior may not be transparent to the pilot resulting in “automation surprises.” Because of this, traditional approaches to training no longer seem adequate to prepare pilots for their new task of supervisory control of highly dynamic, complex systems. These new forms of automation, such as adaptive automation, will require new approaches to and objectives for training.

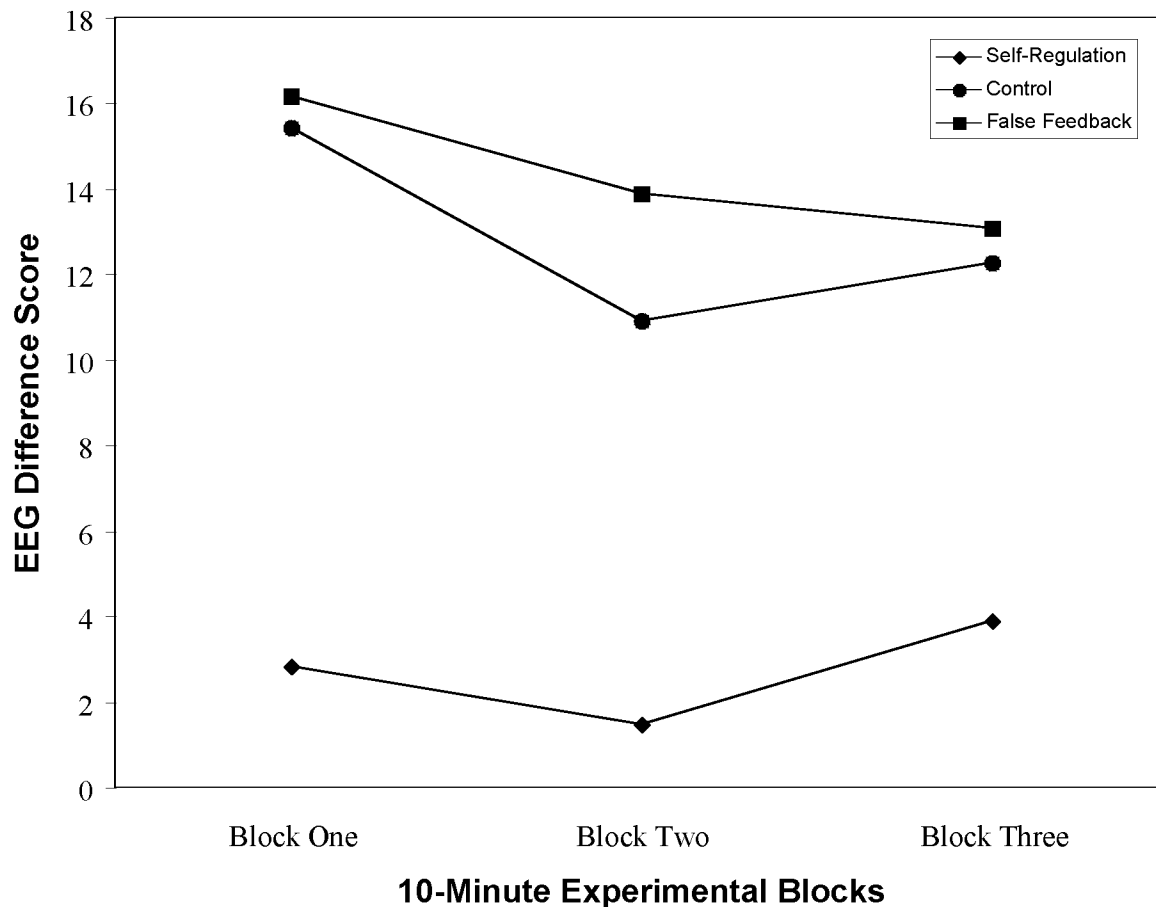
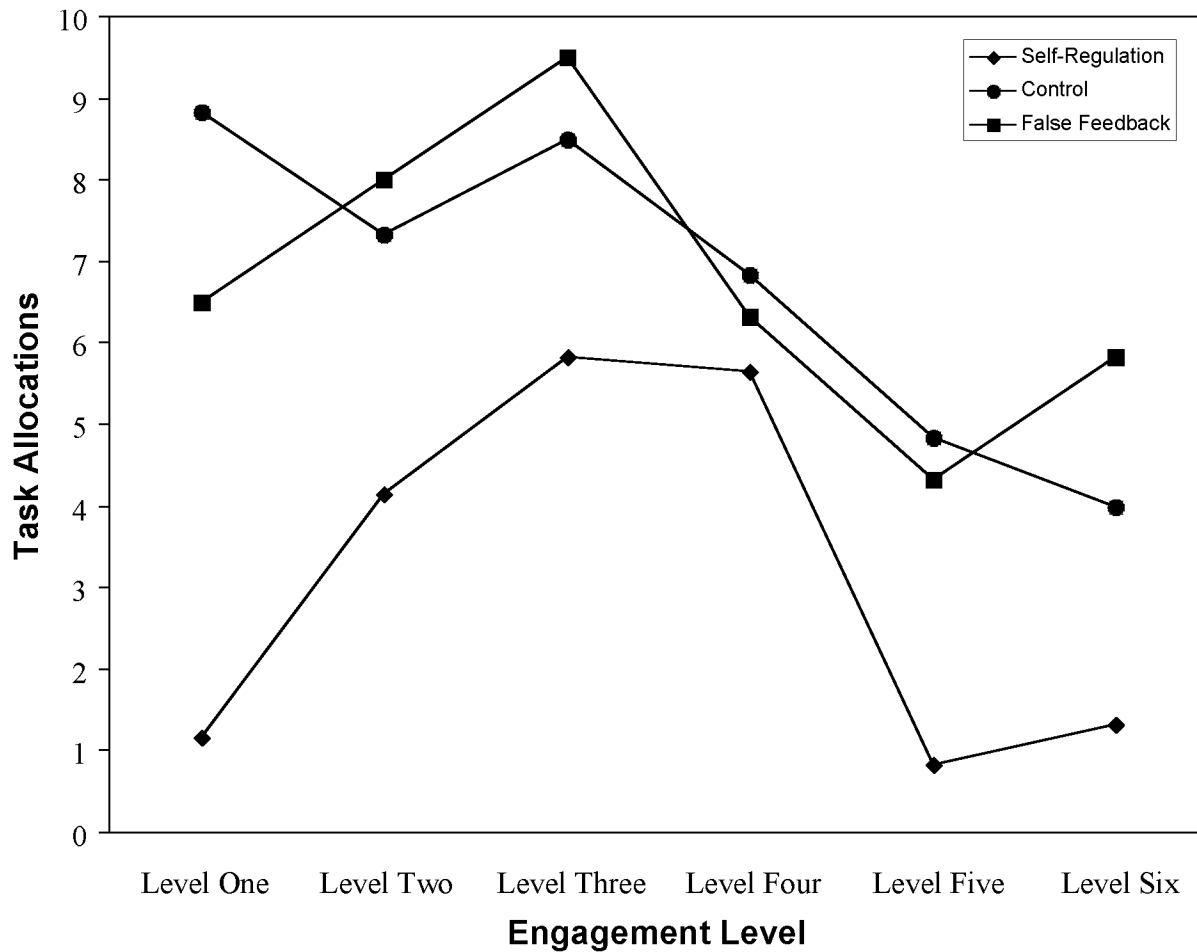


Figure 20. EEG Difference Score Across 10-Minute Experimental Blocks

“Human-centered” automation design details how technology changes human-automation interaction and how best to support the roles that people now have to play as supervisory controllers, exception handlers, and monitors and managers of automated resources (Billings, 1997; Palmer et al., 1994). Self-regulation may represent another tool for supporting human-centered design. Participants in the self-regulation condition were better able to maintain their task engagement level within a narrower range of task modes thereby reducing the need for task allocations. The effect of this was an increase in task performance as well as a decrease in reported workload. Furthermore, these results may have been

due to the increase in return-to-manual performance deficits witnessed in the control and false feedback conditions. Although participants in each condition experienced task allocations, the self-regulation group had fewer task allocations and had significantly lower tracking error scores just after a task allocation than those participants in the false feedback or control conditions. The neurofeedback provided during training may have allowed these participants to better manage their “resources” and thereby regulate their engagement state allowing them to better respond to a change in automation mode. The other conditions,



however, were not given neurofeedback or false feedback and, therefore, the schedule of task allocations may have been opaque to them. Rudisill (1994) reported that many pilots often question “what is it [the automation] doing?” in current pilot-automation interaction (Rudisill, 1994). Therefore, opaqueness is certain to be an important subject of issue with regards to mode unawareness that may develop with adaptive automation. In fact, post-experimental interviews with these participants suggested that they indeed felt unaware as to when and why the task switched from one task mode to another.

Figure 21. Number of Task Allocations Across Adaptive Task Allocation Levels

Scerbo (1996) noted that there is a need to understand how this new form of technology will change the human-automation interaction and to develop training methods to help support the development of adaptive automation. Of course, training cannot and should not be a fix for bad

automation design and there are many issues that still need to be addressed with adaptive automation. Nevertheless, these results support other studies that have demonstrated that physiological self-regulation can help in controlling the onset of hazardous states of awareness, and suggests itself to be a valuable complement to other training procedures for use with adaptive task allocation specifically and intra-personal attention management generally.

Research in the Crew Hazards and Error Management laboratory at NASA Langley Research Center has been directed towards developing a comprehensive strategy for reducing the onset of pilot hazardous states of awareness. It reflects a NASA objective of “making a safe aviation system even safer” by developing methods to dramatically reduce the effects of human error (NASA, 1998; 1999). Our work has focused on a number of areas with the goal of improving cognitive resource management including that of physiological self-regulation. Other areas include adaptive task allocation, adaptive interfaces, hazardous unawareness modeling, cognitive awareness training, and stress-counter response training. These are discussed below. The hope is to design countermeasures and training interventions that may supplement existing strategies, such as crew resource management, but which focuses more on the intra-personal aspects of enhancing flight safety. Together with other NASA-led programs as well as industry and academic partners, the goal of reducing the aircraft accident rate by a factor of 5 within ten years and by a factor of 10 within twenty-five years can become a reality.

### **Cognitive Awareness Training (CATS) Research**

One of the ongoing problems that pilots face today is a diminished state of awareness such as boredom, sleepiness, or fatigue during cruise conditions that could result in various pilot errors. The physiological factors subelement conducted in-house research that utilized a cognitive training exercise to sharpen the pilot’s awareness during simulated flight thereby providing them with a means to overcome these diminished states of awareness. This study utilizes psychophysiological methods in an attempt to assess a pilot’s state of awareness more directly. In turn, the pilots will be able to train themselves to recognize these states of awareness and be more mentally sharp during mundane tasks such as those experienced in cruise conditions. The use of these measurement tools may be beneficial for researchers working to improve aviation safety.

One goal in the study of aviation safety is to try and reduce and possibly eliminate errors caused by poor judgments of the pilot. There are different means currently being employed to help reduce the fatigue and workload of the pilot such as changing some of symbology and visual stimuli that the pilot sees in the flight displays and various other displays on the control panel itself. But of particular interest is the ability to detect these states of awareness so that these cues in the control panel can actually help to reengage the pilot and train them to be more mentally sharp and prepared. The study helps to promote this ideal by utilizing cognitive training exercises that are invoked the moment the data acquisition system recognizes that the pilot is experiencing the aforementioned states of awareness. This study actually is comprised of three phases. The first phase has already been conducted during fiscal year 2000. The second and third phases are planned for FY2001 and FY2002, respectively.

In phase I, a simple flight scenario was developed using Microsoft® flight simulator 2000 where several test subjects demonstrated takeoff, cruise, and aircraft anomaly identification. Reaction time, proper anomaly identification, EEG, and heart rate were the main variables studied in this experiment. The goal is to see if the cognitive exercise has a positive influence on the awareness and the performance of the test subjects. The hypothesis in phase I is that by utilizing these cognitive exercises during cruise flight, the pilot will be more mentally sharp as inferred from a more active EEG signal and will react faster to problems faced in the operations of the flight simulator and also resolve the problems quicker as well. The brain wave patterns will also show a more active state of awareness.

In phase II of the study, test subjects will utilize this same training exercise before flying the test scenario and see if there are any long-term effects of the training. In the third phase of the study, a psychophysiological data acquisition system will be utilized to measure various indices such as EEG or

heart rate real-time to determine when the pilot has indeed begun to experience these diminished states of awareness. Once the system has recognized these diminished states of awareness, the computer will first put the simulated aircraft into autopilot and then invoke this cognitive training exercise by itself to help the test subjects to overcome these previously mentioned states of awareness. Then, the acquisition system will return the pilot back to his normal flight duties. The system will also be tested to help accommodate instances where the test subject has encountered high workload environments and has showed signs of high mental stress. It will then provide some means to take over some of the duties of the flight automatically to help offload and prioritize the tasks for the test subject so that it can reduce and/or manage their level of stress. Once the system detects that the workload has decreased, then the tasks will be returned to the test subject. Again, the goal of this study is to help provide for a new means to deal with these diminished states of awareness that are often experienced by commercial and general aviation pilots.

**Method.** Subjects that were used in phase I of CATS consisted of 12 males whose age range was from 22 to 64 years old. They also possessed 4 to more than 20 years of computer experience using Macintosh®, Microsoft®, and/or UNIX® systems. Their level of education ranged from an Associate of Science Degree to a Ph.D. The attempt was to find test subjects that had some experience using Microsoft® Flight Simulator, but due to the lack of participation from the initial call for test subjects, it was necessary to elicit test subjects who had no experience with this particular software to meet the desired total number of test subjects. As for experience dealing with physiological monitoring, almost 60% of the test subjects have had some experience being monitored, mostly by heart rate sensors.

The test subjects were broken into three groups using a mixed-subjects approach. The three groups consisted of a control group (N=4), a vigilance group (N=4), and an experimental group (N=4). Each group experienced three experimental sessions in which a flying scenario was invoked that included a different aircraft anomaly per session that was preset within the software. The anomalies included failures first, in the altimeter, second, in the attitude indicator (artificial horizon), and lastly, in the vertical speed indicator. Each test subject was required to identify the anomaly when it occurred through a verbal response while at the same time press a button to help time sync the response with the physiological data. They were also given various pre-recorded Air Traffic Control (ATC) commands that provided them altitude and heading information. Prior to the study, the test subjects were given several background questionnaires, which included a biographical questionnaire, Levenson's Locus of Control Scale, Proneness to Boredom Scale, and the Epworth Sleepiness Scale. The data collected from these questionnaires were analyzed prior to the study to help determine a fair and proper distribution of the test subjects amongst the three groups. All the data were then standardized (z scores) to provide equal weight amongst all the scores except that a weighting factor was then added to the flight simulator score experience to enhance its importance in the final score determination. The scores were then tallied into a zsum value. These zsum values for each test subject ranged from -.37 to .81. The test subjects were then assigned sequentially to each group starting with the control group then to the vigilance group then finally to the experimental group using the lowest score first and then building up to the highest score. Again, this procedure was to ensure an even distribution of test subject personalities, abilities, and experience amongst all three groups.

After the proper assignment of test subjects was determined, the test subjects were given assigned dates and times to appear for the simulator. All test subjects were given a 1 hour demonstration of the simulator prior to their assigned test day and a brief description of what was expected of them, but all questions relating to experiment purposes and hypotheses were deferred to the end of their test day. All test subjects were also asked to avoid all caffeine products their assigned day to avoid adding any additional stimulus to their physiological state. The first session on the assigned test day consisted of a short pre-flight briefing, psychophysiological prepping and application, pre-flight questionnaires to provide a subjective means to determine current state of awareness (includes the Stanford Sleepiness Scale<sup>6</sup>, the Terri Dorothy Fatigue Scale<sup>7</sup> (reprinted and modified with permission of the author Terri

Dorothy), and the Cox and McKay Stress Arousal Checklist<sup>8</sup>), takeoff from a pre-determined simulated airport, instructions from ATC to determine required altitude and heading, cruise flight for approximately 24 minutes, then a final heading and direction change with the first anomaly invoked at the 26 minute timeframe, and then once the test subject identified the anomaly, the simulation was paused so that the test subject could fill out the same awareness questionnaires again to determine, subjectively, the test subject's current state of awareness. As for the specific timeframe, it was based on efforts to ensure that vigilance decrement<sup>9</sup> had occurred with the test subjects. After completing the questionnaires, the test subjects were then given an opportunity to experience a landing scenario whose sole purpose was to provide statistical psychophysiological data for other baseline research efforts and was not specifically intended for this experiment. After the test subjects completed the landing scenario, they were then given a 10-minute break.

Upon returning from their break, the test subjects started the same flight scenario again except that two of the groups, the vigilance and the experimental, were also afforded the use of a laptop computer for purposes to be described in further detail below. The control group flew the exact same scenario as before except that the second anomaly was invoked at the 26-minute interval. The difference with the control group and the other two groups was that the vigilance and the experimental groups were given an intervention during flight at the 18-minute time interval and were required to perform their specific tasks for 5 minutes. At the end of the 5 minutes, these test subjects then reengaged the flight scenario and the same anomaly that the control group received at the 26-minute time interval was then invoked and these two groups were asked to properly identify the anomaly as before. Again pre-test and post-test questionnaires were given to each test subject to subjectively determine current state of awareness. The tasks that were given to the vigilance group and the experimental group were vastly different. The experimental group was given a software program known as Captain's Log™ (©1996 Joseph A. Sandford, Ph.D. All Rights Reserved). The original intent of the software is that it was developed as a cognitive training system to help those suffering with Attention Deficit Hyperactivity Disorder (ADHD) and people who have suffered brain maladies such as stroke to perform various exercises that help retrain and refocus their mental abilities. The hope of this experiment was that this same software could be used to help stimulate the test subject's cognitive thinking abilities to sharpen their respective mental state of awareness. The vigilance group was given a mundane and non-stimulating computer vigilance task developed by Dr. Mark Scerbo from Old Dominion University. The purpose of having the vigilance group is to remove any novelty effects that the Captain's Log™ software might produce. That is, the vigilance group is introduced to show that there is hopefully, no effect of stimulation on the test subject due to having a "new" computer task to perform. The hopes are that the Captain's Log™ software in it of itself will produce the necessary cognitive brain stimulation necessary to sharpen the experimental group's state of awareness. After the test subjects completed their respective sessions, the EEG cap and heart rate electrodes were removed from the test subjects and then they were excused for a long lunch break. Again, before the test subjects left the lab, they were instructed to avoid any caffeine products so that no external psychophysiological stimulation was given to them.

Upon the test subject's return, the EEG cap and heart rate electrodes were reapplied and the test subject returned to the simulator. This last session was utilized as a repeat control as a comparison for the first session and to see if there were any carryover effects for the experimental group. The flight scenario was exactly the same as the first session except that a different anomaly was introduced. Again, the test subjects were administered pre-test and post-test questionnaires to gage their relative state of awareness. At the end of this final session, the psychophysiological sensors were completely removed from the test subjects and any excess prepping gel or sensor residue was removed. The test subjects were then provided with a complete description of the experiment in the debriefing session. They were given information on the driving factors that helped to produce the experimental hypotheses along with a description of what was being observed during each session. After this download of information regarding the experiment was given, the test subjects then completed a debriefing questionnaire. The

questionnaire helped to validate the scenarios and sensations experienced by the test subjects along with providing the experimenter with useful information regarding the sensor applications and other environmental lab concerns. The test subjects were then given an opportunity to provide for any useful suggestions, comments, concerns, and/or questions that they might have.

**Results.** The general linear model for repeated measures and the one-way Analysis of Variance (ANOVA) were employed to analyze the data. Currently, the psychophysiological data is still being analyzed. As for the subjective questionnaires, no significant differences were found in the analyses. The only measure of significance that was seen in the analysis occurred in the anomaly reaction time for the different sessions across the groups. A trend leading to a potential significance in the data was observed. According to the analysis, the response time was near significance between groups I and III ( $F(2,9)=4.157, p = .053$ ). When looking at the Tukey and Duncan Post Hoc Tests, significant differences were found between Groups I and Group III for the reaction time response measure. This may indicate that the experimental Group III experienced some effects of the CATS intervention over the control Group I which had no intervention.

**Conclusions and Future Research.** Several factors contributed to the lack of significant results in the data analyzed to date. One in particular was the observation that there were several test subjects who had little or no experience flying the simulator, and they spent a good portion of their time trying to learn how to use the controls and the simulator program. Therefore, their level of engagement and awareness were relatively high throughout the duration of the experiment. This higher level of engagement helped to skew the data. Also, the sample size was probably not large enough to overcome subject to subject variability. But the fact that there is a trend in the reaction time data shows that the results are promising even for a small sample size. It is highly recommended that a further study be conducted with a larger number of test subjects. Also, it is suggested that each test subject be given at least an hour of practice prior to the study to help avoid the learning curve effects on their respective states of awareness. These effects are currently being explored during Phase II of the CATs study.

## Stress-Counter Response Training Research

Increased sophistication in automated aircraft control systems, with multiple backups and automated emergency responses, has steadily increased aircraft reliability. However, despite this or because of this, human error remains the significant cause and limiting reliability factor in aviation incidents and accidents. Ergov (1982) defined the aviator's professional reliability as the ability to handle flight task demands satisfactorily in limited time and solve any problems in an emergency condition. Zhang et al. (1997) suggested that this professional reliability depends on two relative factors: (1) task demand load, and (2) the pilot's cognitive functional capacity. There is considerable evidence to conclude that pilots may and do lose control of their aircraft as a direct result of reactive stress. Foller et al. (1993) noted that such conditions as high task demand and diminished cognitive functional capacity can lead to a narrowing of the focus of attention (i.e., autonomous mode behavior; Li, Shi, & Zhou, 1991; Simonov, Frolov, & Ivanov, 1980) as well as a loss of situation awareness (Endsley, 1994; 1998).

Disasters, such as Three Mile Island, Chernobyl, and USS Vincennes underscore the importance of developing training interventions to offset the real-world stressors on complex cognitive tasks (Johnston & Bowers, 1997). To date, there has been few such training interventions developed and little research to examine how they may improve dynamic decision-making in such stressful environs. Most of these studies have focused on clinical or sports psychology domains and there remains a gap in research in the aviation operational context. The following study will examine a new approach to training adaptive cognitive, autonomic, and behavioral stress responses, termed "Stress Counter Response Training." The approach relies on proven stress exposure training methodologies (SET; Meichenbaum, 1985) and includes task-specific stressors which have been shown to significantly improve performance (Larsson, 1987; Meichenbaum, 1985; Novaco, 1988; Siegel et al., 1981). Furthermore, Stress Counter Response



Training offers the advantages of low time and cost implementation as well as ease of assimilation in already established training interventions (e.g., LOFT, MOST, CRM). Although Crew Resource Management (CRM) training has as an end goal the enhancement of crew decision making in stressful task situations, the problems of human error cannot be addressed by existing CRM training alone. A primary assumption of CRM is that crews will overlearn responses and thereby increase the probability that it will be used during high taskload or emergency situations. However, the tasks of communication, resource management, and coordination become peripheralized during such emergency situations and the pilot's primary focus will be on stick and rudder activities. Therefore, intra-personal CRM, in addition to and in combination with inter-personal CRM training can have significant effects on pilots' ability to effectively deal with emergency situations (Simmons, 1999, Prinzl, Pope, & Freeman, 1999). Therefore, Stress-Counter Response Training has been developed as a valuable adjunct to CRM.

Stress-Counter Response Training is a methodology for training pilots to maintain physiological equilibrium suited for optimal cognitive and motor performance under emergency events in an airplane cockpit. The use of physiology is based on Hockey's (1997) generalized control model that provides mechanisms for dynamic regulatory activity underlying adaptive physiological responses to environmental demands, such as overload, external distraction, and stress. The training method to be tested is novel in that it (a) adapts biofeedback methodology to train physiological balance during simulated operations of an airplane, (b) uses graded impairment of control over the flight task to encourage the pilot to gain mastery over his/her autonomic functions, and (c) can be incorporated into line-oriented flight training (LOFT) or mission-oriented simulation training (MOST) scenarios and substantially improve their effectiveness. The use of a PC-based simulator was based on Baker et al. (1993) who demonstrated the efficacy of PC-based flight simulations using LOFT scenarios for crew resource management training.

**Method.** The research has currently occupied two phases. The first phase is reported here and the second phase has completed experimental runs of subjects, but the data has not been completed analyzed to date. The first study examined the training concept and compared it to more traditional biofeedback methods of stress exposure training as well as a control group that receives no training. Subjects performed mission tasks (MOST-type scenarios) on a PC-based F-15 flight simulator. Measures collected included EKG, SAGAT-based queries, performance measures, verbal protocol analysis, subjective situation awareness measures, and various personality scale measures. The study took place over three intra-experimental phases (Johnson & Cannon-Bowers, 1997) over four experimental sessions.

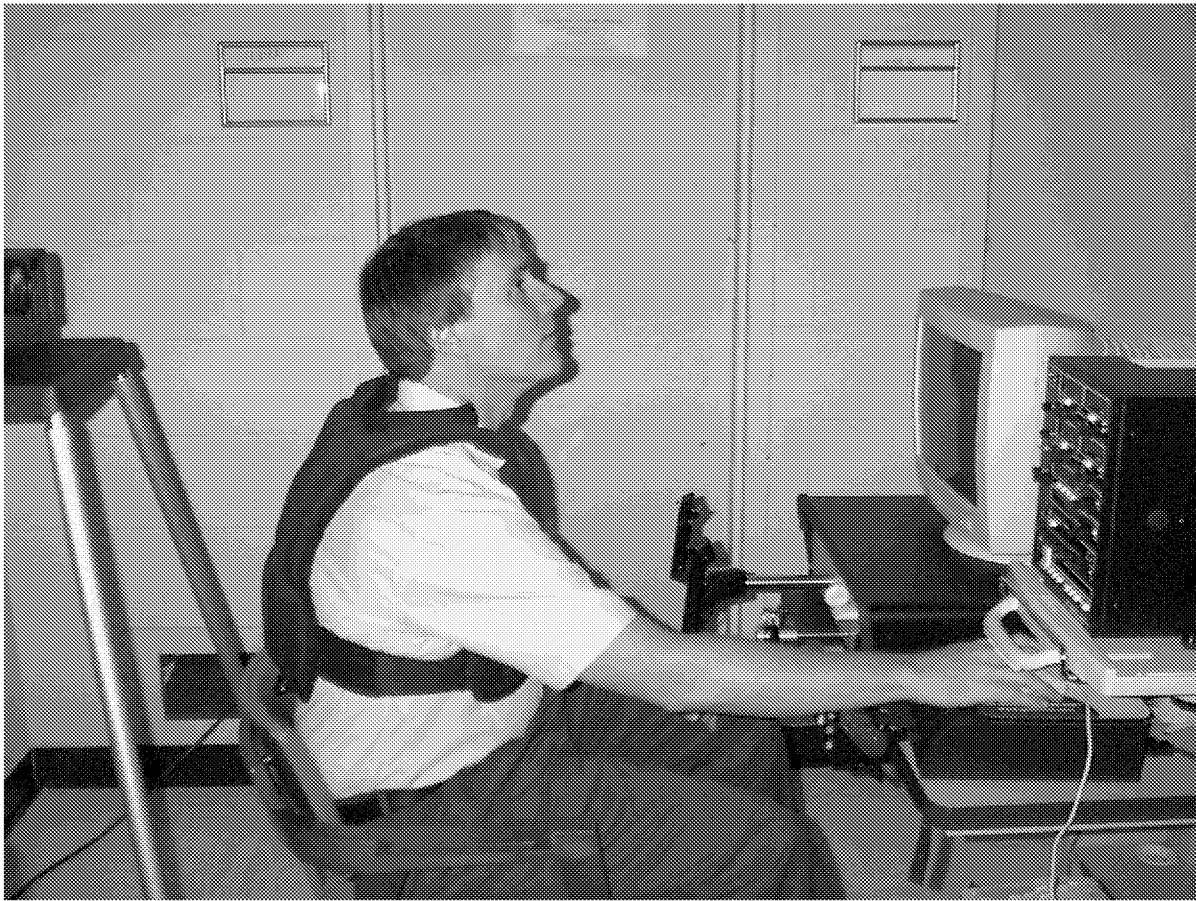


Figure 22. Subject Participant in Stress-Counter Research Experiment

The subjects were 30 females and males who performed a modified PC-based fighter game. The feedback works on the principle of instrumental functionality feedback (IFF) that is embedded into the task to provide adaptive feedback to the subjects by imposing graded impairment of flight controls as their psychophysiology evinces stress responses. The physiological measures we used in the first phase were hand temperature and skin conductance. There were two groups: experimental and control group. The experimental group received graded impairment (through modifications to the control range input of the joystick) when the finger temperature deviated 1 degree Fahrenheit and skin conductance levels of 1 micromho (lost weapon control). The control group received 2-minute reduction of functionality at pre-set times independent of physiology.

**Results.** Unfortunately, the results showed no performance difference between the two groups. However, there was a trend suggesting a steeper level of performance improvement of flight task mission success for the experimental group (see Figure 23).

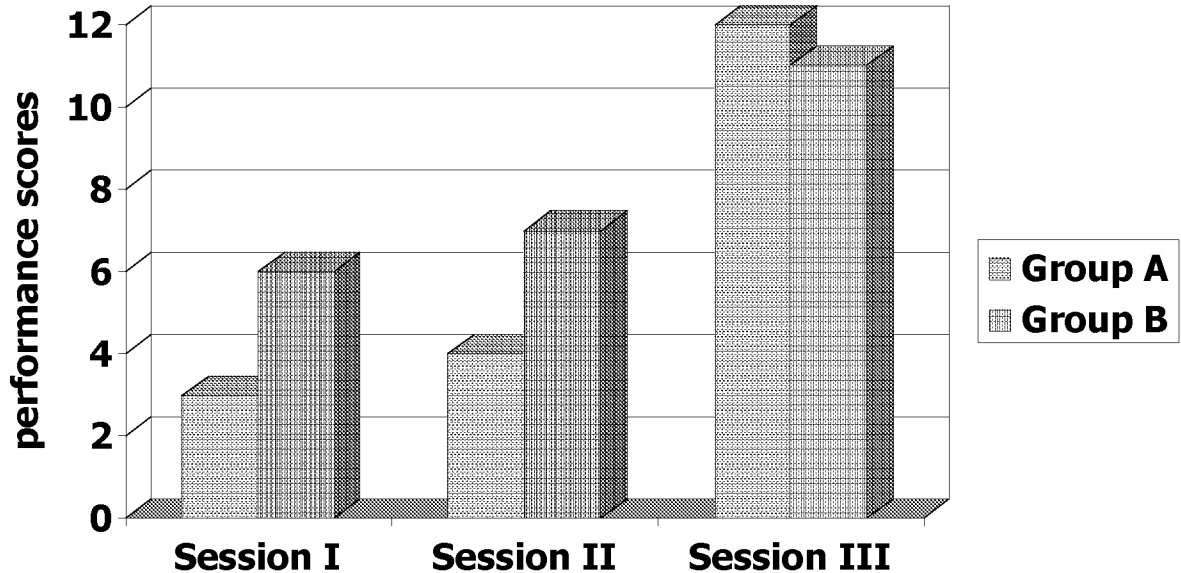


Figure 23. Mean Flight Session Performance in 3 Training Sessions Across Groups.

**Conclusions.** The lack of significant training effects may be due to how we implemented the IFF methodology. We identified several issues after running the experiment as would be expected with such exploratory research including the fact that we imposed unlimited IFF functionality impairment which led to many subjects experiencing unrecoverable loss of control. We also were concerned that the use of hand temperature and skin conductance was not as sensitive a measure of stress as say the EKG or heart-rate. Therefore, we conducted a second experiment that did not impose unlimited impairment and made graduated impaired control inputs based on heart-rate measures. The preliminary results are that the stress-counter response methodology can significantly reduce physiological responses and improve pilot performance under conditions of high workload and high stress. We are very encouraged by these results and are confident that, as they mature, can be a valuable complement to current inter-personal crew resource management techniques that focus on the team element. The techniques of stress-counter response and others, such as CATs and self-regulation, would focus on the intra-personal side that we term, “cognitive resource management” whose acronym is also CRM. Together, they would form a comprehensive training approach to the problem of hazardous states of awareness management in aviation; a training approach we call CRM<sup>2</sup>.

## SPIN-OFF RESEARCH

NASA Langley Research Center continues to develop new technologies to address flight deck human factors issues with psychophysiological methods, while seeking opportunities to transfer the technologies into educational and clinical applications. Research has developed technologies using physiological measures for assessing pilot stress, sustained attention, engagement and awareness in a laboratory flight simulation environment. Biocybernetic systems employing these measurements have been designed to be used for evaluating manned system designs for compatibility with human capabilities. Biomedical spin-offs have emerged from this work through collaboration with medical centers.

## Videogame Neurofeedback

Neurofeedback (NFB) training systems provide real-time information to trainees showing them how well they are producing the brainwave patterns that research has shown to distinguish normal individuals from individuals classified as ADHD. Neurofeedback training can be a long and arduous procedure, creating adherence and attrition problems. An engaging form of training delivery is needed. An entertaining way of delivering neurofeedback training evolved from the physiological factors research conducted at NASA Langley Research Center (U.S. Patent No. 5,377,100). Brainwave-assisted videogames are designed to respond to electrical brain activity as well as gamepad or joystick input. In the current experimental embodiment, mastery of off-the-shelf videogames depends on proper cognitive engagement, reflected in a high Engagement Index, as well as good game skill. This innovative approach to neurofeedback, which is based on patented NASA technology, will potentially allow subtle training of enhanced capacity for concentration while individuals enjoy playing their favorite videogames, and may make it possible for neurofeedback training to become an integral part of video- game home entertainment for children and adults. There are a number of modifications being made to the patent and, therefore, specifics as to the operation of the videogame neurofeedback technology cannot be described in technical details here. A diagram of the operation of the videogame neurofeedback technology is presented in Figures 24 and 25.

**Method.** A study was conducted to research whether the technology has any positive effect beyond that of traditional biofeedback in the treatment of ADHD children. 22 children with ADHD of the hyperactive impulsive subtype (DSM-IV criteria plus physician diagnosis) were the subjects. They were between ages 9-13 years and 3 were girls and 19 were boys. All the children were on short-acting medications for ADHD and had normal intelligence and no history of affective problems or learning disabilities.

The children were randomized to treatment groups of videogame or standard neurofeedback. Children in both groups completed 40 individual treatment sessions, once to twice per week or 20-40 weeks of sessions. The children came for one test session before and after treatment while they completed quantitative EEG tests, tests of variables of attention (TOVA), and neuropsychological tests. BASC monitor data and actigraph (physical activity) data was collected pre- and post-treatment and every ten sessions. Children in both groups were trained with a single active Cz electrode, with reference electrode and ground attached to earlobes.

The videogame group played standard Playstation console games (Spyro the Dragon, Tony Hawk, and Gran Turismo) and the neurofeedback was embedded within the games. Training consisted of fixed-length training intervals interspersed with listening and reading tasks. The standard control group received neurofeedback through the Thought Technology Procomp+ hardware and Multitrace software package. Displays were bar graphs and simple figures representing changes in somato-motor rhythms and beta and theta EEG bands. Like the videogame group, training consisted of fixed-length training intervals interspersed with listening, reading, and unmodulated videogame playing. The control group training resembles in every way, the traditional and typical neurofeedback training applied to the treatment of ADHD.

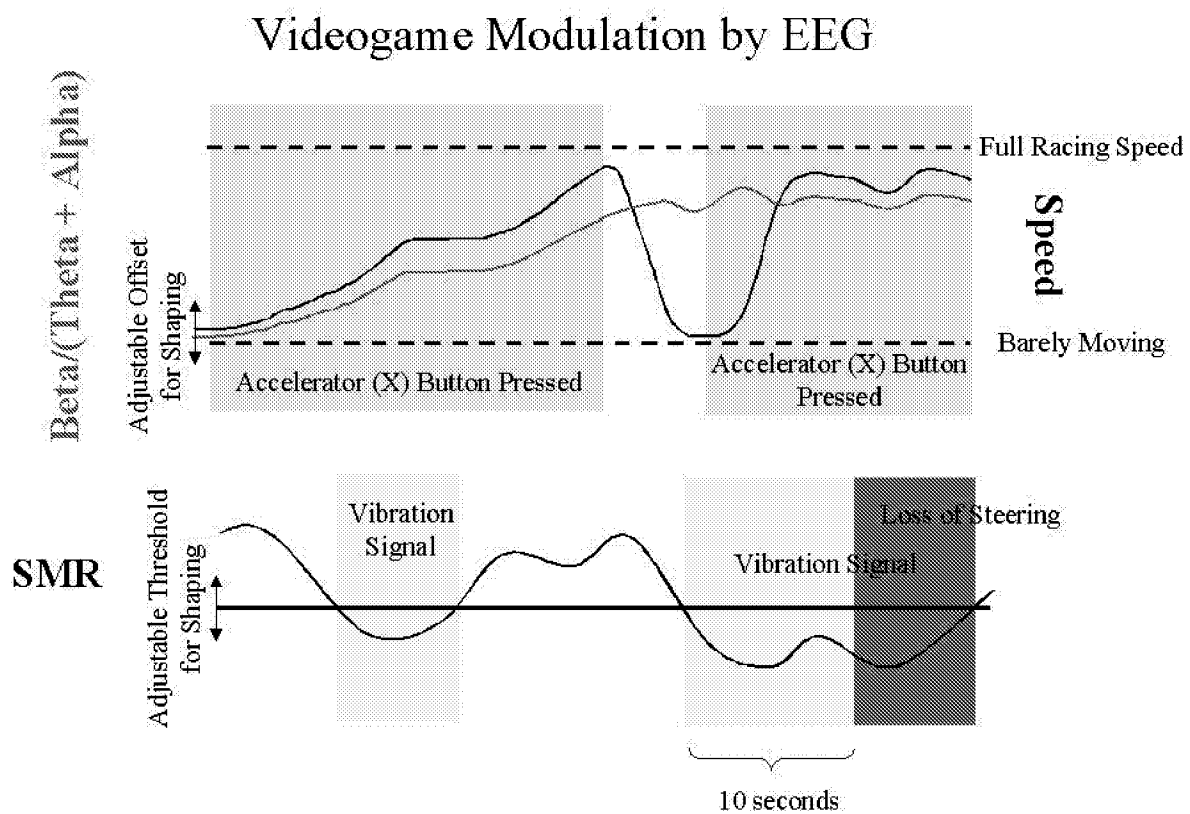


Figure 24. Video-Game Control Modulation

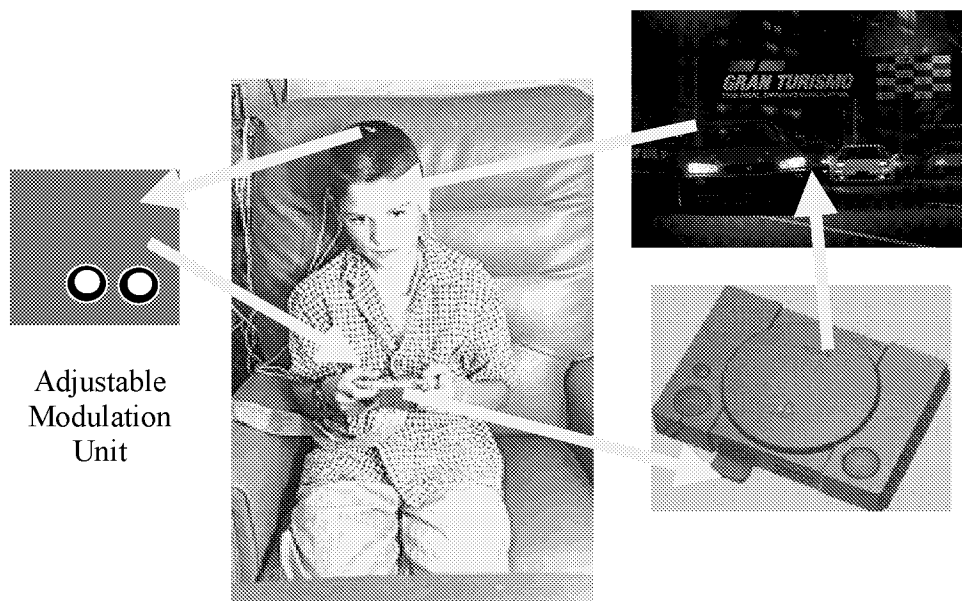


Figure 25. Videogame Group Set-up

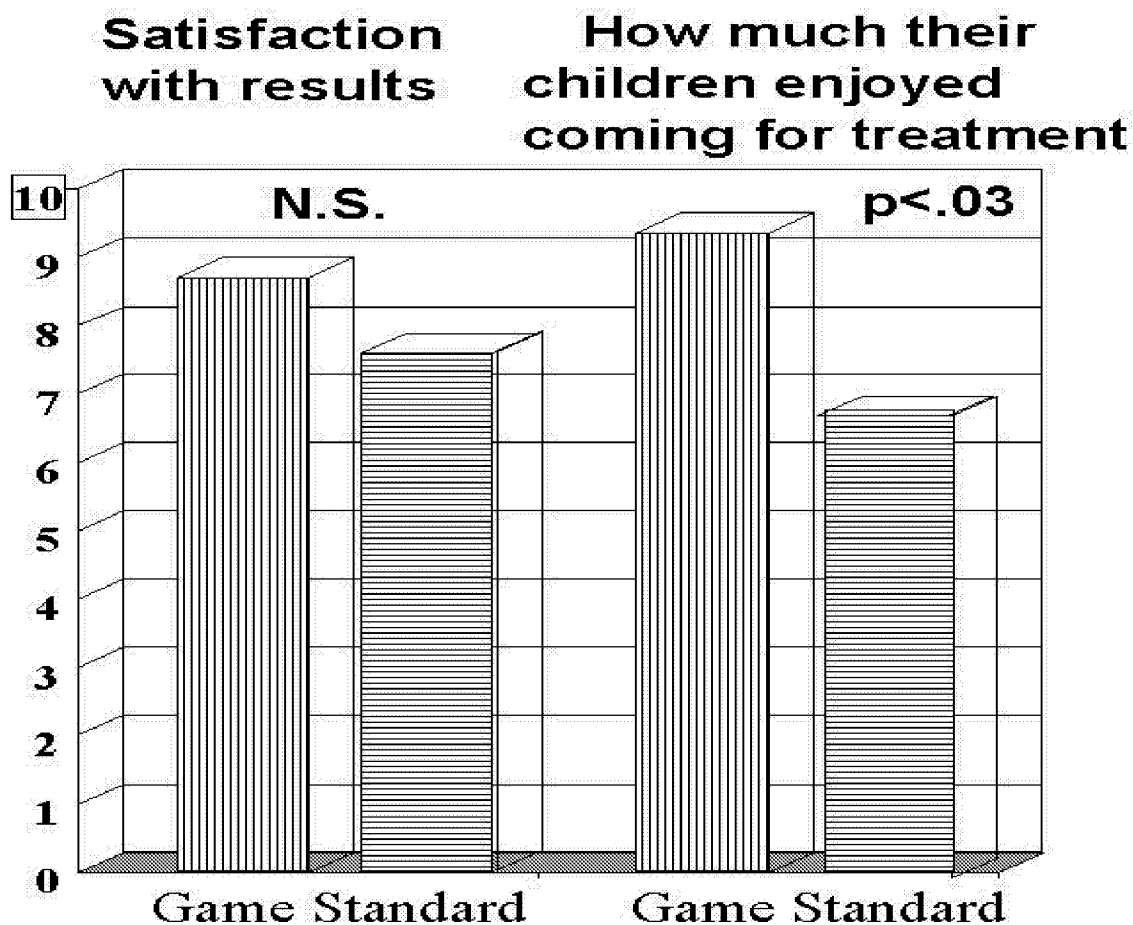


Figure 26. Satisfaction Results of Study

The results of the study showed no significant differences between the traditional and the videogame technology groups. Therefore, this would suggest that the videogame technology and protocol performed comparably with proven and well-accepted practice of biofeedback treatment of ADHD. This was exactly what we hoped for. However, what was even more remarkable was the one significant finding that we found in terms of motivation to continue treatment. There was a significant difference between those subjects who got the videogame treatment and those who got the control treatment. One of the largest problems with the treatment of ADHD children is keeping them motivated to continue throughout 40 or more sessions. In fact, most children do not complete treatment and instead have to rely on medications to control their ADHD. These results suggest that the problem of non-motivation can be dealt with effectively through the application of one of the favorite pastimes of children, playing games that has embedded ADHD biofeedback treatment. Clinical trials are continuing at present to further investigate the potential of this technology to substantially change the treatment approach that is taken with ADHD children and adults. We believe that the videogame biofeedback technology has a number of advantages: Being inherently motivating, blending of different treatment approaches, allowing individuals

to select the games that they like best and allowing for ease of update for remaining current with the videogame available at present, providing for cross-gender treatment of ADHD since the child can select what game they like and therefore are not limited to mostly male-oriented games, and it can be used at home without clinical intervention saving time and money.

## **VISCEREAL Diabetes Technology**

VISCEREAL is a virtual reality system that non-invasively renders physiological information in such a way as to accurately represent, in real-time and on-line, the functioning of the underlying physiological sources in both the appearance and action (LaRC Patent Case No. LAR-15396-P). In the field of psychophysiology, biofeedback closed the loop by providing patients with real-time information about the functioning of their own physiology-information previously observed by the clinician-to help patients learn physiological self-regulation.

The purpose of VISCEREAL is to immerse a patient in a real-time display environment that facilitates learning about physiological function as well as learning of voluntary control of function. It also immerses a physician in a real-time display environment that facilitates visualization of physiological function for diagnosis as well as monitoring of drug response. VISCEREAL stands for VISCERal + REALity = VISCEREAL; it is non-invasive endoscopic biological feedback. The current embodiment that is subject to patenting uses general relaxation techniques and provides temperature and blood volume feedback for the treatment of Raynaud's disease, migraine headaches, vasoconstriction secondary to diabetes and connective tissue disease, and hypertension. More than 15 million Americans who have diabetes may soon use NASA virtual reality technology as a new treatment in the self-management of the disease. Preliminary observations show that NASA's artificial-vision technology can help patients at risk for nerve damage associated with diabetes to visualize and control blood flow to their arms and legs. This application, which comes from several years of research aimed at enhancing aviation safety, combines two technologies: sensors to measure the body's reactions and powerful computer graphics to turn those measurements into a 3-D virtual environment (see Figure 27). The graphics technologies are used in research with cockpit artificial-vision systems to help pilots seen low- or no-visibility situations, and as data-visualization tools to help designers study air-flow patterns around new aircraft shapes, as well as adaptive automation technology developed at NASA Langley Research Center. Using biofeedback methods the patients will increase blood flow, which will be measured through sensors attached to their fingertips. The system uses skin-surface pulse and temperature measurements to create a computer-generated image of what is actually happening to blood vessels under the skin. Just as pilots use artificial vision to "see" into bad weather, patients will use this virtual reality device to see beneath their skin.

We have engaged the Strelitz Diabetes Research Institutes of the Eastern Virginia Medical School (EVMS) to conduct clinical trials with the new technology. Furthermore, trials are also underway at the Behavioral Medicine Center at the University of Virginia Health Sciences Center to evaluate the technology for treatment of other blood-flow disorders.

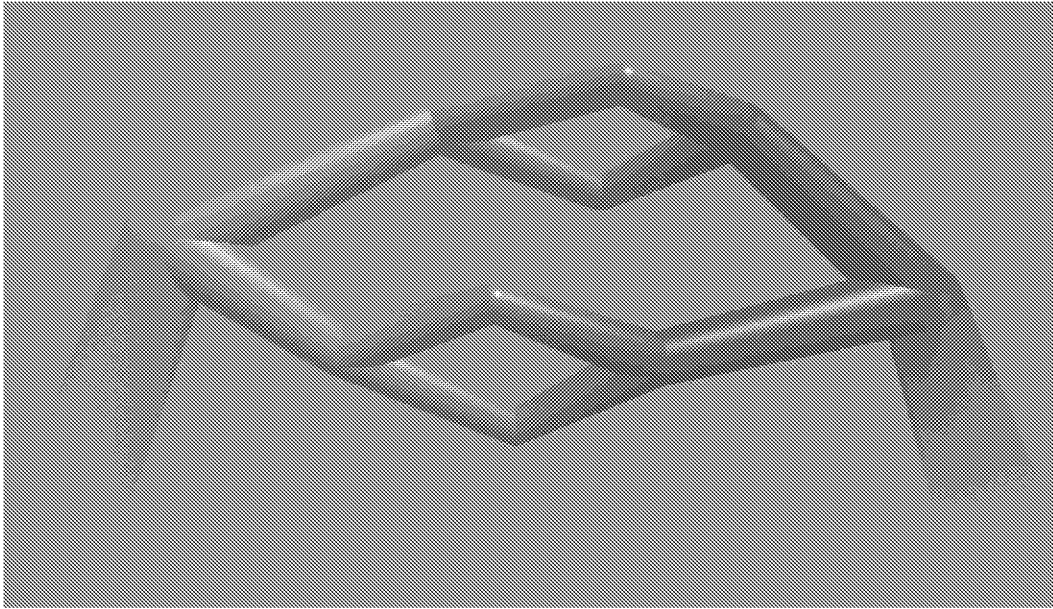


Figure 27. The VISCEREAL Display

## **Crew Response Evaluation Window**

CREW is a human response measurement technology useful for pharmaceutical testing, product usability testing, and medical research. The Crew Response Evaluation Window (CREW) technology permits the evaluator to select and simultaneously view several, previously dispersed, sources of physiological and behavioral response information in a single, integrated display window (LaRC Patent Case No. LAR 15367-1).

NASA LaRC researchers developed the Crew Response Evaluation Window (CREW) technology to improve the process of monitoring the responses of pilots in flight research experiments. CREW can also be used to evaluate the effects of pharmaceuticals, products and medical disorders on human behavior. NextGen Systems, Inc., the company to which the technology has been licensed, plans to also use the technology for objective broadcasting and advertisement analysis. The CREW technology has been licensed to NextGen Systems, Inc. (Blue Bell, Pa.). A subsidiary group has emerged that focuses just on this technology, called Capita Systems. NextGen Systems, Inc., designs and markets systems and services that measure psychological engagement, receptiveness, and communication effectiveness. These systems utilize electroencephalogram (EEG) and the CREW technology licensed under an exclusive agreement from NASA to measure electrical activity in the human brain. The technology was developed from research on the biocybernetic software system as a method to evaluate automated flight deck concepts for compatibility with human capabilities. Our research has focused on development of the methods to determine the optimum mix of allocated human and automated tasks in a cockpit. Since licensing the technology from NASA, NextGen has engaged in significant research and development to further refine the suitability of the original NASA software and position it for use in media testing.



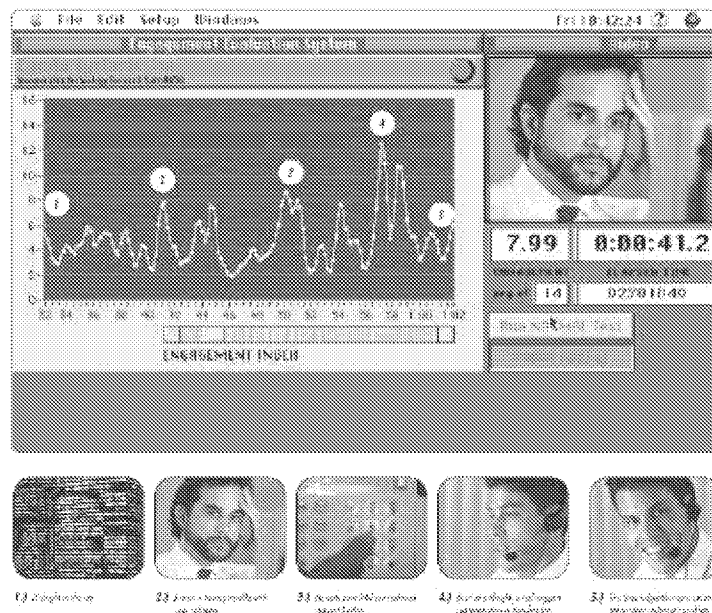


Figure 28. Capita Systems Adaptation of CREW Technology for Marketing Purposes

Advertising research has long recognized the need to develop and place commercial messages that maintain a viewer's attention, interest or involvement. With the fragmentation of traditional demographics, the proliferation of special interest groups joined globally through the Internet, and the growth of affinity marketing, the need for advertisers to optimize their media dollars through appropriate content, context and placement has never been more acute. Research to determine the efficacy of advertising and commercial placements has grown into a multibillion industry. As an outgrowth of its preliminary marketing efforts, NextGen has developed relationships with a number of prominent media, entertainment, and marketing industry leaders.

The work with CREW has resulted in revenues “spinning back” to NASA and has been adopted, in addition to the ADHD and VISCEREAL technologies, by the NASA commercialization office. The CREW technology has been provided with U.S. Copyright registration (TXU743936) and is currently undergoing patent review (08/641,041). To date, Capita Systems has performed test media services (e.g., assessment of commercials) for 17 companies, including MTV and Fortune 100 companies, and two incumbents for congress.

## CURRENT RESEARCH FOCUS

### Real-Time Adaptive Automation of EICAS

A cooperative agreement with Catholic University will be expanding on previous research that used a model-based approach to adaptive automation to control the automation modes in a simulated Engine Indicator and Crew Alerting System (EICAS) display. The EICAS is the standard engine and system health monitoring system used in many advanced glass-cockpit aircraft. The previous study assigned 24 rated pilots to three groups: a workload-matched adaptive group, a “clumsy automation” group, and a control group. A 60-min session was used comprising three phases of high-low-high task load to simulate a profile of takeoff/climb, cruise, and approach/landing phases of flight. For the

workload-matched group, adaptive automation in the form of automation of the vertical dimension of the two-dimensional tracking task on the MAT was used during the high task load phases at the beginning and end of the session. There was also used a temporary return of the automated EICAS task to manual control during the middle of the second, low taskload phase. For the clumsy-automation group, these contingencies were reversed, so that they received aiding when taskload was low and were allocated the automated EICAS task when taskload was high. Pilots have often complained that automation increases workload at times when workload is already high (e.g., going “heads-down” and reprogramming the FMS during descent and approach phases) and reduces workload when it is not necessary (e.g., cruise phase of flight). For the control group, adaptive automation was not provided. Overall performance on the three tasks of the MAT was better for the workload-matched group suggesting that a model-based approach to adaptive automation has potential for improving human-automation interaction and modulating workload demands in the performance of flight tasks. Furthermore, it was found that supervisory control behavior was significantly improved for the workload-matched group and worse for the clumsy-automation group providing further suggestions that adaptive automation, if implemented correctly, can potentially reduce monitoring errors of lapses, slips, and mistakes and thereby reduce automation-induced complacency.

The present focus of these studies is on the real-time analog of adjusting task mode in the EICAS automation based on psychophysiological measures of heart-rate variability combined with pilot performance. The rationale for this set of studies is that the model-based approach presumes to be able to accurately gauge that pilot workload is indeed high during postulated periods of high workload. For example, the descent phase of flight has been reported to be higher in workload than cruise phases of flight. However, not always or not to the extent that it would be significant to need intervention from the technology that adaptive automation could provide. Rather, it would be best to be able to monitor in real-time pilot workload and make those adjustments when it is clear that it is needed. This would account for the stark individual differences that have been well documented in how pilots perceive and react to workload.

The first study will use the same experimental design reported above with the exception that adaptive automation will be implemented, instead of based on a script, in real-time using a moving window of heart-rate data. The EICAS task will be under automation control throughout the 90-min session except that it will be returned to manual control at random intervals. All 30 subjects are told that the automation is not 100% reliable and should be monitored for failures. Adaptive automation will be implemented through the task allocation of the lateral control part of the tracking task. The design and choices made in the implementation of adaptive automation were based on a pilot study of 5 pilots. The decision to make task mode adjustments (i.e., adaptive automation) will be based on a 5-min window of heart-rate data that is then updated every 10 seconds. After a minimum of 5 minutes has elapsed into the flight task, the estimates of workload will be compared to the transition points determined in the pilot study. At each point in time  $t$ , the moving window estimate of a parameter  $A_t$  will be compared to its corresponding transition point value. If the estimate falls out of the range of the transition points established for that parameter, the adaptive logic will be triggered for the tracking task. For the EICAS task, the moving window estimate will be of correct percent detection of malfunctions. In this case, if the estimate falls below the desired set point, the adaptive logic will be triggered and the EICAS system will be controlled to manual control in 30 seconds and then returned to automated control after a period of 5 minutes.

There will be three groups of 10 assigned to the adaptive automation, clumsy automation, or control groups (see above for descriptions). Dependent measures include the performance on the tasks, detection rate on the EICAS task, and workload measures from the .1 Hz heart-rate variability and NASA-TLX. A liberal algorithm will be used to make the task allocation changes using an “OR” triggering algorithm in which either the tracking RMSE or .1 Hz measure of heart-rate variability can trigger the change in lateral control of tracking. The work is exploratory in that it will hopefully lead to the development of “AND” algorithms or a more complex logic involving non-linear combination of parameters (e.g., with the use of a neural network model).

## Information-Processing Stages and Adaptive Automation

The capabilities and potential benefits of automation of complex system operations have often been overestimated and oversold by technologists. In considering aviation systems, including aircraft and air traffic control (ATC) workstations, the current role of automation is limited because of the limitations of expert systems (Leroux, 1993). In general, automation is not capable of higher-order cognitive functions, such as information integration and decision making, which are required in piloting tasks and ATC operations for effective performance (Leroux, 1993). Humans must remain part of decision making processes in the control of such system in order to ensure performance. The key limitation of automation for ATC is the lack of capability of, for example, an expert system to consider the context of a decision and to quickly select an alternative, as humans often do on the basis of decision making heuristics and biases.

With this in mind, some researchers (c.f., Laois & Giannacourou, 1995) have posed the question as to whether automation should only be applied to, for example, data acquisition and communication tasks, in the context of aviation system operations, versus it being applied to decision making functions or tasks requiring higher-order aspects of information processing. Laois and Giannacourou (1995) stated that automation is generally better for monitoring tasks whereas humans are better at decision making, especially in critical situations. That is, automation is most suited to early sensory and information acquisition stages of information processing while humans are well-suited to the latter (i.e., advanced stages of processing). They studied human performance in an ATC simulation and surveyed expert controllers to determine the implications of automation of ATC decision making functions on performance. They observed significant performance decrements when futuristic forms of automation (conflict projection and clearance advisory) were applied to decision functions in the simulation, particularly when high-level automation was used. The survey results indicated that automation should only be applied to data acquisition and communication versus conflict projection and clearance advisory.

This research suggests that caution should be exercised when considering the application of automation to aviation systems because of limitations in current technology and the implications of automation on human operator performance when applied to advanced functions, such as decision making. Laois & Giannacourou results demonstrate that automation of certain ATC functions may not support the overall objective of automation - to augment operator skills.

Adaptive automation (AA) or dynamic function allocation (DFA) has been explored as a potential solution to automation capability issues and the documented negative effects of full automation on human operator performance, including operator complacency, vigilance decrements, and loss of situation awareness (SA) over short periods of time and skill decay over long durations. Unfortunately, current AA literature only broaches the central issue discussed above; that is, the human performance implications of automation of complex systems functions requiring higher order aspects of information processing. Kaber (1997), Kaber and Riley (1999) and Parasuraman et al. (2000) all reviewed a number of empirical studies of AA that have focused on the performance effects of dynamic function allocation in complex systems, specifically monitoring and psychomotor functions (e.g., tracking). This literature has also pointed to the limited number of studies that have investigated the implications of AA on cognitive task performance (e.g., Hillburn et al., 1997). It is important to note that in such research cognitive functions or tasks may not have been passed between a human operator and computer, but, rather, the human and computer may exchange perceptual and psychomotor functions and the effects of the exchange on human cognition are evaluated.

Some work has indirectly investigated the implications of AA of lower-order aspects of human-machine system information processing and pointed to the need to study [AA] of the advanced stages of information processing in complex systems operations, including decision making and response execution. Crocoll and Coury (1990) evaluated the human performance consequences of automation

reliability when applied to information acquisition and analysis as part of human-machine system performance. This work is relevant to the current proposal, as AA may be considered a form of unreliable automation. That is, depending upon the state of a system and its task, the automation may be turned “on” or “off”. This may or may not occur with operator notification. In the latter case, the operator may in fact perceive AA as unreliable automation. Crocoll and Coury’s (1990) work attempted to define the conditions under which automation reliability does or does not affect human performance. They compared information acquisition/analysis automation with decision automation. Research has shown that people can adapt to automation unreliability when automation is applied to low-level information processing functions. It has also been suggested that negative effects of automation unreliability may be more pronounced for decision automation compared to information analysis automation (Parasuraman et al., 2000). Only one study has looked at this issue and Parasuraman et al. (2000) have pointed to the need to further examine whether automation unreliability has greater negative effects on the advanced stages of human-machine system information processing than monitoring and information analysis.

Research is needed at this point, to describe human responses to AA of complex system functions requiring higher-order aspects of information processing, and to establish the ability of humans to adapt to AA of such functions in comparison to their ability to adapt to automation unreliability (failures) when automation is applied to early sensory and information acquisition functions of complex systems. Manning (1993) described a general three-step procedure for evaluating effects of complex system automation on human operator performance (i.e., conducting the type of research which has been identified as being necessary). The steps included: (1) identifying the objective of the automation (what is it to do); (2) identifying the client of the automation; and (3) defining the needs of the client (their information requirements). The final step is also aimed at discovering whether automation will prevent in anyway users from acquiring needed information, and identifying the level of human involvement in system operations necessary to prevent complacency and vigilance decrements, and to maintain SA. This information can be used to determine whether automation is meeting the identified objective and whether it will ultimately augment human operator skill. This approach is similar to contemporary methods for design of human-automation interaction in complex systems. Parasuraman et al. (2000) formulated a model-based approach to automation of complex systems (e.g., ATC systems) based on existing theories of human information processing. The model included seven steps: (1) identify system functions to be automated; (2) identify the type of automation (the stage of information processing to which automation is to be applied); (3) identify the degree of automation; (4) evaluate the human performance consequences of applying automation; (5) make initial specification of types and levels of automation; (6) evaluate the reliability of the automation and associated costs; and (7) final specification of automation. Four stages of human-machine system information processing are considered in this model, including information acquisition, information analysis, decision-making and action, to describe the degree of automation for the operation of a complex system. These stages correspond to aspects of human information processing included in historical models (e.g., Broadbent, 1958), such as perception, planning, decision-making, and action.

Parasuraman et al. (2000) approach to automation design and Manning’s approach to evaluating automation can be used to characterize various types of human-machine systems in terms of the aspects of information processing required for effective performance. They may also serve to categorize the functions of human-machine systems in terms of operator information requirements and stages of information processing. Therefore, the approaches could be used to identify functions requiring higher-order cognition and facilitate examination of the application of AA to such functions and evaluation of the affect on human performance. In general, these methods to automation design and evaluation need to be evaluated through future AA research and fieldwork.

We have established a cooperative agreement with Dr. David Kaber of North Carolina University to determine the specific performance and workload effects of AA of information acquisition, information analysis, decision-making and action functions as part of complex system performance. That is, the research will seek to quantify, for example, the exact effect of AA of the decision making function of the

Multitask© simulation on overall human-machine system performance. In general, this goal will be achieved by applying a method to automation evaluation similar to Manning's (1993) general approach.

The main hypothesis of this work is that humans will not be able to adapt to AA of decision making and action functions, as part of complex system performance, as well as they are able to use AA of information acquisition and analysis functions. Furthermore, it is speculated that application of AA to the decision making aspect of performance will not be as effective as AA of the monitoring or information analysis aspects of the task for managing operator workload. Human monitoring performance relies on the short-term sensory stores and perception, including detection and pattern recognition. Decision making relies not only on the perceptual process but information storage in memory, integration of information on perceived stimuli and long-term memory (LTM) structures in working memory, development of a situation model, classification of the situation model in terms of schema and scripts in LTM, and use of decision making heuristics and biases for response selection. A potential problem with AA of a decision making function is that there are many stages of information pre-processing that humans undertake in order to make decisions and periodically removing and/or involving a person in a complex system control loop may be disruptive to the cognitive processes critical to a decision. In monitoring, the cognitive pre-processing is limited in comparison to decision making.

With respect to the hypothesis on workload management through AA, the critical difference between monitoring function performance and decision-making is that the former usually doesn't require information storage, or the signal-response (S-R) associations for monitoring are usually automatic in comparison to S-R associations in complex decision-making. Therefore, in monitoring humans are not usually required to recall (from time-to-time) information in LTM in order to keep track of system states. Complex decision making tasks differ from choice-reaction tasks in that they usually occur over extended periods of time and require significant recall and integration of information. It is expected that the application of AA to decision making as part of complex system operations might cause operators to attempt to retain system information in WM or LTM from one manual control period to another. This is, however, not expected to be the case in monitoring function performance. Therefore, the effectiveness of AA for managing human operator workload, when applied to the decision making aspect of human-machine system performance, may be limited in comparison to the effectiveness of AA for the same purpose when applied to a monitoring function. These hypotheses will be investigated through the designed experiment.

The experiment to be conducted as part of this research will require subjects to perform in a dual-task paradigm. Not only will they control the Multitask© simulation (a flight simulation suite that was developed under the direction of Dr. David Kaber), but also they will perform a secondary task, a gauge-monitoring task. The experimental scenario will involve subjects acting as air defense system radar operators on-board an AWACS aircraft (Boeing 707), who also have a responsibility to monitor the status of aircraft subsystems, for example engine temperature or oil pressure. The gauge-monitoring task will present a fixed-scale display with a moving pointer and subjects will be required to monitor and maintain the pointer within a designated "acceptable" region on the display. They will accomplish this by using keys on a keyboard to facilitate corrective motion of the pointer if it deviates into "unacceptable" regions at either end of the fixed-scale display.

The gauge-monitoring task is psychomotor in nature involving subjects monitoring, diagnosis and action. Subjects will be instructed during the experiment to focus their attention on the Multitask© and to allocate remaining attentional resources to performance of the secondary task. During the experiment, the secondary task will be used as an objective measure of subject workload due to the Multitask©. Performance in the gauge monitoring task will be recorded in terms of the ratio of the number of unacceptable pointer deviations detected by subjects to the number of deviations simulated (i.e., the "hit-to-signal" ratio).

Criteria will be established for performance of the secondary task as part of the dual-task scenario. The criteria will be based on subject gauge-monitoring performance in a pilot study. The

performance criteria will also be criteria of primary task workload. A criterion for task underload and overload will be established and associated with Multitask© adaptive function allocation (interface changes). Task overload will be indicated by significant performance degradations in the secondary task, representing excessive levels of primary task workload, and will be associated with mandates for subjects to activate automation in the primary task (Multitask©). Task underload will be indicated by near-perfect performance in the gauge-monitoring task and will be associated with mandates for subjects to deactivate automation of the primary task function and perform the Multitask© manually. Since the perceptual and cognitive demands of Multitask© functions completely overlap those of the gauge-monitoring task, previous research (Kaber & Riley, 1999) has found the gauge-monitoring task to be a sensitive indicator of workload changes in the Multitask© simulation, as affected by AA.

The independent variable to be controlled in this experiment will be the stage of information processing to which AA is applied. Therefore, there will be four levels of the variable, including information acquisition, information analysis, decision making, and action implementation as part of Multitask© performance.

With respect to response measures, Manning (1993), in evaluating ATC automation, measured human-machine system performance in terms of productivity, or the number of flights handled by a controller. He also measured automated system functioning in terms of reliability. With this in mind, performance in the Multitask© will be measured in terms of the number of targets eliminated, the number of targets overlooked/missed, and the number of target collisions. This will allow for a performance-based assessment of AA as applied to the different aspects of information processing as part of the complex system control.

Beyond performance measures, subjective workload assessments will also be made during the study using the Modified-Cooper Harper (MCH) scale. This measure focuses on mental workload caused by interface design. Observations on the measure will be used to verify objective measures or workload using the secondary task.

A between-subjects design will be used in order to minimize the potential for Multitask© training carry-over effects from one experimental trial to another. Four groups of subjects will be formed on the basis of the Multitask© function to which AA is applied. One group, will experience AA applied to the information acquisition stage of information processing as part of Multitask© (i.e., monitoring targets). Depending upon the observed level of subject workload, motion of the portal display will either be controlled by the subject or the computer system. A second group will be exposed to AA applied to the information analysis stage of information processing as part of Multitask© performance (i.e., analyzing target characteristics and behavior). Depending upon subject workload, automation may be provided in the form of a summary display on target characteristics and potential conflicts. The third subject group will be exposed to AA of the Multitask© decision-making function. If subject workload is high (secondary task performance is poor), they will be provided with the target elimination advisory aid. The aid will provide instructions as to which targets on the display to eliminate and when. The fourth group will experience AA of the response execution aspect of Multitask performance. Under automated control, the computer will formulate a target elimination plan and eliminate targets on the basis of the plan. Manual control of the action function will require subjects to process targets using the computer keyboard or mouse.

For comparison purposes, two control conditions will also be studied as part of the experiment. An additional group of subjects will be recruited to perform the Multitask© simulation without any adaptive interface aids being activated. These subjects will also be required to perform the secondary task in order to ensure a fair comparison of overall human-machine system performance across the AA and completely manual control conditions of the experiment. The second control condition will involve full automation of all functions as part of Multitask© operation. This condition will be investigated to establish the maximum performance capability of the automation. No subjects will be used in evaluation

of this condition, as there will be no role for the human to fulfill. The computer will process targets based on an algorithm considering the speed of targets, their positions, etc.

## **Biocybernetic Human “In-The-Loop” Studies**

Research will continue in-house and with a cooperative agreement with Old Dominion University (Dr. Fred Freeman, Dr. Mark Scerbo, and Dr. Peter Mikulka) and a Graduate Student Research Fellowship (GSRP) to the University of North Carolina (awarded to graduate student, Michael Clamann) that focuses on the continued testing and refinement of adaptive automation algorithms using psychophysiological measures and the closed-loop, biocybernetic system developed at the NASA Langley Research Center. Current research will focus on the supervisory control tasks, vigilance tasks, feature-integration tasks, and PC-based Microsoft Flight Simulation tasks. Also, a new approach is being developed that will use a combination of EEG, performance, and heart-rate measures to make task allocation decisions. The details of the operation of the system were described previously in this section and therefore will not be repeated here.

One set of research will employ a multi-task paradigm to compare and contrast adaptive automation strategies based on physiological and secondary task measures of workload. Specifically, subjects will be required to perform a dynamic control task integrated with a signal detection task. An index of user arousal will be computed based on EEG signals and hit-to-signal ratios will be calculated on signal detection task performance in order to predict automated and manual control allocations of the dynamic control task. The frequency and duration of allocations, when using each type of trigger, will be compared. In addition, measures of dynamic control task performance will be recorded. This information will be used to directly validate the control strategies defined using the triggers, and indirectly validate the physiological and secondary task measures of workload as triggers. It is also expected that the research will identify a superior trigger for various levels of difficulty of the dynamic control task manipulated during the research. A second set of studies will be similar to this research but will use a Triesman feature-integration paradigm that requires cognitive rather than perceptual modes of information processing.

Other research will examine the combination of EEG, performance, and heart-rate measures as the “trigger” for adaptive automation. The research will be conducted in two phases. The first phase will consist of a pilot study in which baseline data will be collected. Subjects will perform tasks from the MAT while their EEG and heart rate are being monitored. After being given 30 minutes of training on the tasks, subjects will be run at three levels of difficulty: low, medium, and high difficulty. Data from this pilot study will be analyzed to determine:

- Which combination of EEG sites should be used in construction of an EEG engagement index for use in an adaptive automation system?
- Should the EEG index just be used for adaptive task allocation when it reflects a low level of arousal?
- How well does heart rate correlate with performance on the low and high difficulty task modes?
- Do subjective measures such as the NASA-TLX correlate with the physiological measures?

In Phase 2, the data from Phase 1 will be used in the construction of an adaptive automation system that employs multiple psychophysiological measures to drive the system. Subjects will then be tested using the new system on the MAT initially and more ecologically valid flight tasks in later research studies.

## **Spin-Off Research**

NASA Langley Research Center will continue to explore spin-off applications of the research. Dr. Lance Prinzel and Dr. Alan Pope have recently submitted a patent application that applies our experience in adaptive automation, biofeedback, and sports psychology. The technology will help athletes, such as tennis players and golfers, improve stress and anxiety coping skills based on some principles and technologies we have developed in our past research. Additional details cannot be made available at present because of intellectual property as directed by the NASA Langley Research Center Commercialization and Patents Office.

In addition to our spin-off research with athletes, we will also continue our research and development of technology in the treatment of ADHD, blood-flow disorders, and anxiety / stress disorders. In 2000, the technology for the treatment of ADHD children was voted as #9 of NASA's top 10 overall innovations for the year. The work was recognized because of its potential to significantly improve the quality of life for millions of children that suffer from ADHD. Improving quality of life is one of NASA's primary objectives and mission statements and, therefore, we are hopeful and encouraged to be able to contribute to such a worthwhile goal.

## **NASA Program Collaborations**

Researchers, involved in the research described in the technical memorandum, have worked in collaboration with other NASA aviation programs, such as synthetic vision systems (SVS). Our work in the field of psychophysiology, human performance assessment, human factors, and psychology has been utilized to help with many of the same issues that other programs are currently facing. Issues, such as task overload, cognitive capture, loss of situation awareness, etc., are important areas of concern for many other aviation research programs. And, our research in these areas and the assessment tools that we have developed have been leveraged by these programs to help ensure that these new technologies being developed in these programs address these issues. Currently, we are part of a research team examining low visibility, loss-of-control issues involved in low hour general aviation accidents. Our team is measuring psychophysiological responses and gathering stress and arousal measures when pilots enter into loss-of-control situations. The work will serve both the program of SVS general aviation element but also base human factors research by helping to uncover the etiologies and precipitating factors that are contributors to entering into these hazardous states of awareness.



## REFERENCES

- Arnegard, R. (1991). Operator strategies under varying conditions of workload. Ph.D. thesis, Old Dominion University.
- Barber, B. (1983). Logic and the Limits of Trust. New Brunswick, NJ: Rutgers University Press.
- Billings, C.E., (1997). Aviation automation: The search for a human-centered approach. Mahwah, NJ: Lawrence Erlbaum Associates.
- Billings, C., Lauber, J., Funkhouser, H., Lyman, G., & Huff, E. (1976). NASA Aviation Safety Reporting System (NASA TM-X-3445). Moffett Field, CA: NASA Ames Research Center
- Byrne, E. A. & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological Psychology, 42, 249-26
- Becker, A. B., Warm, J. S., Dember, W. N., & Hancock, P. A. (1991). Effects of feedback on workload in visual performance. In Proceedings of the Human Factors and Ergonomics Society 35<sup>th</sup> Annual Meeting (pp. 1491-1494). Santa Monica, CA: Human Factors and Ergonomics Society.
- Billings, C. E. (1997). Aviation Automation: The Search for a Human-Centered Approach. New Jersey: Lawrence Erlbaum Associates.
- Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. British Journal of Clinical Psychology, 21, 1-16.
- Byers, J.C., Bittner, A.C., & Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), Advances in Industrial Ergonomics and Safety I. London: Taylor & Francis.
- Byrne, E.A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological Psychology, 42, 249-268.
- Comstock, J. R., & Arnegard, R. J. (1992). The Multi-Attribute Task Battery for human operator workload and strategic behavior research (NASA TM-104174). Hampton, VA: NASA Langley Research Center.
- Cox, D.J., Kovatchev, B.P., Morris, J.B., Phillips, C., Hill, R.J., & Merkel, L. (1998). Electroencephalographic and psychometric differences between boys with and without attention-deficit/hyperactivity disorder (ADHD): A pilot study. Applied psychophysiology and biofeedback, 23, 179-188.
- Cunningham, S. C., and Freeman, F. G., (1994). The electrocortical correlates of fluctuating states of attention during vigilance tasks. (NASA Research Grant NCC1-176 Old Dominion University).
- Damos, D.L., Smist, T.E., & Bittner, A.C. (1983). Individual differences in multiple-task performance as a function of response strategy. Human Factors, 25, 215-226.
- Davidson, R.J. (1988). EEG measures of cerebral asymmetry: Conceptual and methodological issues. International Journal of Neuroscience, 39, 71-89.

- Davies, D. R., & Parasuraman, R. (1982). The psychology of vigilance. London: Academic Press.
- Deaton, J. E., & Parasuraman, R. (1993). Sensory and cognitive vigilance: Effects of age on performance and subjective workload. Human Performance, 6, 71-97.
- Dittmar, M. L., Warm, J. S., Dember, W. N., & Ricks, D. F. (1993). Sex differences in vigilance performance and perceived workload. Journal of General Psychology, 22, 239-245.
- Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 41-62). North-Holland: Elsevier Science Publishers.
- Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. Human Factors, 37, 381-394.
- Farmer, R., & Sundberg, N. D. (1986). Boredom proneness- the development and correlates of a new scale. Journal of Personality Assessment, 50, 4-17.
- Federal Aviation Administration (1990). The national plan for aviation human factors: Volume 1. Washington, D.C.
- Freeman, F. G., Mikulka, P. J., Scerbo, M. and Hadley, G. (1998). Comparison of two methods for use in adaptive automation. Perceptual and Motor Skills, 86, 1185-1186.
- Freeman, F.G., Mikulka, P.M., Scerbo, M.W., Prinzel, L.J., & Clouatre, K. (in press). Evaluation of a psychophysiological controlled adaptive automation system using performance on a tracking task. Applied Psychophysiology and Biofeedback.
- Freeman, F.G., Mikulka, P.J., Prinzel, L.J., & Scerbo, M.W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking system. Biological Psychology, 50, 61-76.
- Fulop, A., & Scerbo, M. W. (1991, September). The effects of event rate on perceived workload and boredom in vigilance. Poster presented at the Human Factors Society 35<sup>th</sup> Annual Meeting, San Francisco, CA.
- Galinsky, T. L., Dember, W. N., & Warm, J. S. (1989, March). Effects of event rate on subjective workload in vigilance. Poster presented at the Human Factors Society 35<sup>th</sup> Annual Meeting, San Francisco, CA.
- Grubb, P. L., Miller, L. C., Nelson, W. T., Warm, J. S., & Dember, W. N. (1994). Cognitive Failure and Perceived Workload in Vigilance Performance. Proceedings of the Automation and Human Performance Conference, 1, 115-121.
- Hancock, P.A., & Warm, J. S. (1989). A dynamic model of stress in sustained attention. Human Factors, 31, 519-537.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139-183). Amsterdam: North-Holland.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. Cognitive Sciences, 3, 275-310.

Kaber, D. B. & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on workload assessment through a secondary monitoring task. In Scerbo, M. W. and Mouloua, M. (Eds.) Automation technology and human performance: Current research and trends. (pp. 129-133). Mahwah NJ: Erlbaum.

Kellar, M.A., Folen, R.A., Cowings, P.S., Toscano, W.B., & Hisert, G.L. (1993). Autogenic-feedback training improves pilot performance during emergency flying conditions (NASA TM-104005). Moffett Field, CA: Ames Research Center.

Langer, E. (1989). Mindfulness. Reading, MA: Addison-Wesley.

Lee, J. D. (1992). Trust, self-confidence, and adaptation to automation. Unpublished Doctoral Thesis. University of Illinois, Urbana, IL.

Lee, J., & Moray, N. (1992). Trust and the allocation of function in the control of automatic systems. Ergonomics, 35, 1243-1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40, 153-184.

Lubar, J.F. (1991). Discourse on the development of EEG diagnostics and biofeedback for Attention-Deficit/Hyperactivity Disorders. Biofeedback and Self-Regulation, 16, 201-225.

Maurino, D.E., Reason, J., Johnston, N., & Lee, R.B. (1995). Beyond aviation human factors: Safety in high technology systems. Brookfield, VT: Ashgate.

Montgomery, D. A. & Gronert, M. E. (1999) Effects of selected task performance criteria at initiating adaptive task reallocations. In Proceedings of the Human Factors and Ergonomics Society 43<sup>rd</sup> Annual Meeting. (pp. 1389) Houston, TX

Morrison, J.G., & Gluckman, J.P. (1994). Definitions and prospective guidelines for the application of adaptive automation. In M.Mouloua & R.Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 256-263). Hillsdale, NJ: Lawrence Erlbaum, Associates.

Mosier, K.L., Skitka, L.J., & Korte, K.J. (1994). Cognitive and social psychological issues in flight crew/ automation interaction. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: current research and trends (pp. 256-263). Hillsdale, NJ: Earlbaum.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. International Journal of Man-Machine Studies, 27, 527-539.

Muir, B. M. (1989). Operators' trust and use of automation controllers in a supervisory process control task. Ph.D. Thesis, University of Toronto, Canada.

Muir, B. M. (1994). Trust in automation. Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics, 37, 1905-1922.

Muir, B. M. & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, 39 (3), 429-460.

Mulholland, T. (1977). Biofeedback as a scientific method. In G. Schwartz & J. Beatty (Eds.), Biofeedback: Theory and research (pp. 9-28). New York: Academic Press.

National Aeronautics and Space Administration (1998). NASA Policy Directive (NPD)-1000.1. Washington, D.C.

National Aeronautics and Space Administration (1999). Roadmaps to the Future: Strategic Roadmaps in Support of the Three Pillars and Ten Goals. Washington, D.C.

Norris, S.L., & Currier, M. (1999). Performance enhancement training through neurofeedback. In J.R. Evans & A. Abarbanel (Eds.), Introduction to quantitative EEG and neurofeedback (pp. 224-239). New York, NJ: Academic Press.

Palmer, M.T., Rogers, W.H., Press, H.N., Latorella, K.A., & Abbott, T.S. (1994). A crew-centered flight deck design philosophy for high-speed civil transport (HSCT) aircraft (NASA Technical Memorandum 109171). Hampton, VA: Langley Research Center.

Parasuraman, R. (1990). Event-related brain potentials and human factors research. In J.W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), Event-related brain potentials: Basic issues and applications (pp. 279-300). New York: Oxford University Press.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". International Journal of Aviation Psychology, 3 (1), 1-23.

Parasuraman, R., Mouloua, M. and Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. Human Factors, 38, 665-679.

Parsons, H.M. (1985). Automation and the individual: Comprehensive and comparative views. Human Factors, 27, 99-112.

Pope, A. T. & Bogart, E. H. (1992). Identification of hazardous awareness states in monitoring environments. SAE Technical Paper #9211369, SAE 1992 Transactions: Journal of Aerospace, Sec. 1-REM 101 pp.449-457.

Pope, A.T., Bogart, E.H., & Bartolome, D.S (1995). Biocybernetic system validates index of operator engagement in automated task. Biological Psychology, 40, 187-195.

Prinzel, L.J., Freeman, F.G., Scerbo, M.W., & Mikulka, P.J. (in press). A biocybernetic system for adaptive automation. International Journal of Aviation Psychology.

Prinzel III, L.J., Hadley, G., Freeman, F.G., & Mikulka, P.J. (1999). Behavioral, subjective, and psychophysiological correlates of various schedules of short-cycle automation. In M.Scerbo & K.Krahl (Eds.), Automation technology and human performance: Current research and trends. Mahwah, NJ: Lawrence Erlbaum Associates.

Prinzel III, L.J., Scerbo, M.W., Freeman, F.G., & Mikulka, P.J. (1997). Behavioral and physiological correlates of a bio-cybernetic, closed-loop system for adaptive automation. In M. Mouloua & J.M. Koonce (Eds.), Human-Automation interaction: Research and practice (pp. 66-75). Mahwah, NJ: Lawrence Erlbaum Associates.

Prinzel, L. J. & Freeman, F. G. (1997). Task-specific sex differences in vigilance performance. Subjective workload and boredom. Perceptual & Motor Skills, 85 (3), 1195-1202.

Ray, W., & Cole, H. (1985). EEG alpha activity during cognitive processing: Influence of attentional factors. International Journal of Psychophysiology, 3, 43-48.

Rencken, W.D. & Durrant-Whyte, H. F. (1993). A quantitative model for adaptive task allocation in human-computer interfaces. IEEE Transactions on Systems, Man, and Cybernetics, 23, 1072-1090.

Riley, V. (1989). A general model of mixed-initiative human-machine systems. In Proceeding of the Human Factors and Ergonomics Society 33<sup>rd</sup> Annual Meeting, (pp. 124-128). Denver, CO: Human Factors and Ergonomics Society.

Riley, V. (1994). A theory of operator reliance of automation. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 8-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), Automation Theory and Applications. Human Factors in Transportation, (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates.

Rouse, W.B. (1977). Human-computer interaction in multitask situations. IEEE Transactions Systems, Man, and Cybernetics, SMC-7, 293-300.

Rudisill, M. (1994). Flight crew experience with automation technologies on commercial transport flight decks. In M.Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 203-211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sarter, N.B., & Woods, C.D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of flight management system. International Journal of Aviation Psychology, 4, 1-28.

Sawin, D. A., & Scerbo, M. W. (1994a). Vigilance: How to do it and who should do it. In Proceedings of the Human Factors and Ergonomics Society 38<sup>th</sup> Annual Meeting (pp. 1312-1316). Santa Monica, CA: Human Factors and Ergonomics Society.

Sawin, D. A., & Scerbo, M. W. (1994b). Vigilance: How much is enough? Insight: The Visual Performance Technical Group Newsletter, 16(2), 1-3. (Human Factors and Ergonomics Society).

Sawin, D. A., & Scerbo, M. W. (1995). Effects of instruction type and boredom proneness in vigilance: Implications for boredom and workload. Human Factors, 37 (4), 752-765.

Scerbo, M. W. (1996). Theoretical perspective on adaptive automation. In R. Parasuraman and M. Mouloua (Eds.) Automation and human performance: Theory and applications (pp. 37-63). Mahwah NJ: Erlbaum.

Scerbo, M. W., Greenwald, C. Q., & Sawin, D. A. (1993). The effects of subject-controlled pacing and task type upon sustained attention and subjective workload. Journal of General Psychology, 113, 293-307.

Scerbo, M. W., Rettig, K. M., & Bubb-Lewis, C. L. (1994). A validation study of a task-related boredom scale. In Proceedings of the 2<sup>nd</sup> Mid-Atlantic Human Factors Conference, (pp. 135-136). Washington, D. C.

Schutte, P.C., Latorella, K.A., Comstock, J.R., Rudisill, M., & Trujillo, A. (in press). Countermeasures for skill, rule, and knowledge-based errors. Proceedings of the Eleventh Aviation Psychology Symposium. Columbus, OH: Ohio State University.

Schutte, P.C., & Willshire, K.F. (1997). Designing to control flight crew errors. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 1978-1983.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), Measures in health psychology: A user's portfolio (pp. 35-37). Windsor, UK: NFER-NELSON.

Seamster, T.L., Boehm-Davis, D.A., Holt R.W., & Schultz, K. (1998). Developing advanced crew resource management (ACRM) training: A training manual. Washington, D.C., Federal Aviation Administration.

Sheridan, T. B. & Ferrell, L. (1974). Man-Machine Systems. Cambridge, MA: MIT Press.

Simmon, D.A. (1998). Boeing 757 CFIT Accident at Cali, Columbia, becomes focus of lessons learned. Flight Safety Digest, 17, 1-31.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of a complacency-potential scale. International Journal of Aviation Psychology, 3, 111-122.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Individual differences in monitoring failures of automation. Journal of General Psychology, 120 (3), 357-373.

Singh, I. L., Molloy, R., & Parasuraman, R. (1997). Automation-induced monitoring inefficiency: role of display location. International Journal of Human-Computer Studies, 46, 17-30.

Sirevaag, E., Kramer, A., Wickens, C.D., Reisweber, M., Strayer, D., & Grenell, J. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. Ergonomics, 36, 1121-1140.

Stark, J. M. & Scerbo, M. W. (1998). The effects of complacency potential and boredom proneness on perceived workload and task performance in an automated environment. Paper presented at the Human Factors and Ergonomics Society 42<sup>nd</sup> Annual Meeting. Chicago, IL: Human Factors and Ergonomics Society.

Streitberg, B., Röhm, J., Herrmann, W.M., & Kubicki, S. (1987). COMSTAT rule for vigilance classification based on spontaneous EEG activity. Neuropsychobiology, 17, 105-117.

Thackray, R. I., & Touchstone, R. M. (1989). Detection efficiency on an air traffic task with and without computer aiding. Aviation, Space, and Environmental Medicine, 60, 744-748.

Wickens, C.D., Isreal, J.B., & Donchin, E. (1977). The event related cortical potential as an index of workload. Proceedings of the Human Factors & Ergonomics Society, 21, 282-287.

Wickens, C.D., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychological analysis of the reciprocity of information processing resources. Science, 221, 1080-1082.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). An introduction to human factors engineering. New York: Addison-Wesley Longman. (pp. 392-395)

Wiener, E. L., & Curry, R. E. (1980). Flight deck automation: Promises and problems. Ergonomics, 23, 995-1011.

Wiener, E. L. (1981). Complacency: Is the term useful for air safety? In Proceedings of the 26<sup>th</sup> Corporate Aviation Safety Seminar (pp. 116-125). Denver: Flight Safety Foundation, Inc.

Wiener, E. L. (1989). Human factors of advanced technology ("glass cockpit") transport aircraft (NASA TP-117528). Moffett Field, CA: NASA-Ames Research Center.

Wiener, E.L., Kanki, B.G., & Helmreich, R.L. (1993). Cockpit resource management. New York, NJ: Academic Press.

Will, R. P. (1991). True and false dependence on technology: Evaluation with an expert system. Computers in human behavior, 7 (3), 171

Woods, D.D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R.Parasuraman & M.Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Assoc.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 2002	3. REPORT TYPE AND DATES COVERED Technical Memorandum		
4. TITLE AND SUBTITLE Research on Hazardous States of Awareness and Physiological Factors in Aerospace Operations		5. FUNDING NUMBERS WU 711-50-21-01		
6. AUTHOR(S) Lawrence J. Prinzel III				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-2199		8. PERFORMING ORGANIZATION REPORT NUMBER L-18149		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001		10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA/TM-2002-211444		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified-Unlimited Subject Category 54                      Distribution: Nonstandard Availability: NASA CASI (301) 621-0390		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) The technical memorandum describes research conducted to examine the etiologies and nature of hazardous states of awareness and the psychophysiological factors involved in their onset in aerospace operations. A considerable amount of research has been conducted at NASA that examines psychological and human factors issues that may play a role in aviation safety. The technical memorandum describes some of the research that was conducted between 1998 and 2001, both in-house and as cooperative agreements, which addressed some of these issues. The research was sponsored as part of the physiological factors subelement of the Aviation Operation Systems (AOS) program and Physiological / Psychological Stressors and Factors project. Dr. Lance Prinzel is the Level III subelement lead and can be contacted at l.j.prinzel@larc.nasa.gov.				
14. SUBJECT TERMS Human-Automation Interaction, Psychophysiology, Hazardous States of Awareness, Training			15. NUMBER OF PAGES 88	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	